

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:50:23 ; Search time 394.5 Seconds
(without alignments)
12915.082 Million cell updates/sec

Title: US-09-787-016A-1

Sequence: 1 cccggtgcccgcgcgcacac.....accttaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%
Maximum Match 100%

Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgcn2_1/USPTO.spool.p/US09787016/runat_20042004_154658_17003/app.query.fasta_1.2759
-DB=Pending Patents AA Main -OPMT=fastan -SUFFIX=pct.rapm -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -INITS=bite -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=1000 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000
-USER=US09787016 @CGN 1.1 444 @runat_20042004_154658_17003 -NCPU=6 -ICPU=3
-NO MMAP -LARGEJQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAPN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Pending Patents AA Main:

1: /cgcn2_6/prodata/2/paa/PCTUS_COMB.pep:*
2: /cgcn2_6/prodata/2/paa/US06_COMB.pep:*
3: /cgcn2_6/prodata/2/paa/US07_COMB.pep:*
4: /cgcn2_6/prodata/2/paa/US080_COMB.pep:*
5: /cgcn2_6/prodata/2/paa/US081_COMB.pep:*
6: /cgcn2_6/prodata/2/paa/US082_COMB.pep:*
7: /cgcn2_6/prodata/2/paa/US083_COMB.pep:*
8: /cgcn2_6/prodata/2/paa/US084_COMB.pep:*
9: /cgcn2_6/prodata/2/paa/US085_COMB.pep:*
10: /cgcn2_6/prodata/2/paa/US086_COMB.pep:*
11: /cgcn2_6/prodata/2/paa/US087_COMB.pep:*
12: /cgcn2_6/prodata/2/paa/US088_COMB.pep:*
13: /cgcn2_6/prodata/2/paa/US089_COMB.pep:*
14: /cgcn2_6/prodata/2/paa/US090_COMB.pep:*
15: /cgcn2_6/prodata/2/paa/US091_COMB.pep:*
16: /cgcn2_6/prodata/2/paa/US092_COMB.pep:*
17: /cgcn2_6/prodata/2/paa/US093_COMB.pep:*
18: /cgcn2_6/prodata/2/paa/US094_COMB.pep:*
19: /cgcn2_6/prodata/2/paa/US095_COMB.pep:*
20: /cgcn2_6/prodata/2/paa/US096_COMB.pep:*
21: /cgcn2_6/prodata/2/paa/US097A_COMB.pep:*
22: /cgcn2_6/prodata/2/paa/US097B_COMB.pep:*
23: /cgcn2_6/prodata/2/paa/US098_COMB.pep:*
24: /cgcn2_6/prodata/2/paa/US099A_COMB.pep:*
25: /cgcn2_6/prodata/2/paa/US099B_COMB.pep:*
26: /cgcn2_6/prodata/2/paa/US100_COMB.pep:*
27: /cgcn2_6/prodata/2/paa/US101_COMB.pep:*

28: /cgcn2_6/prodata/2/paa/US102_COMB.pep:*
29: /cgcn2_6/prodata/2/paa/US103_COMB.pep:*
30: /cgcn2_6/prodata/2/paa/US104_COMB.pep:*
31: /cgcn2_6/prodata/2/paa/US106_COMB.pep:*
32: /cgcn2_6/prodata/2/paa/US107_COMB.pep:*
33: /cgcn2_6/prodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
No matches found					

Search completed: April 20, 2004, 21:12:13
Job time : 394.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:51:33 ; Search time 19.5 Seconds
(without alignments)
7386.308 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtgcccgcgcgcac.....actcttaagatcatatccgcg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 196679 seqs, 27592530 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cg2_1/USPTO.spool.p/US09787016/runat.20042004.154659.17029/app.query.fasta_1.2759
-DB=Pending Patents AA New -QFMT=fastran -SUFFIX=pct.rapn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=1000 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1.1.62 @runat.20042004.154659.17029 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending Patents AA New:*
1: /cg2_6/prodata/1/paa/PCT_NEW_COMB.pep:*
2: /cg2_6/prodata/1/paa/US06_NEW_COMB.pep:*
3: /cg2_6/prodata/1/paa/US07_NEW_COMB.pep:*
4: /cg2_6/prodata/1/paa/US08_NEW_COMB.pep:*
5: /cg2_6/prodata/1/paa/US09_NEW_COMB.pep:*
6: /cg2_6/prodata/1/paa/US10_NEW_COMB.pep:*
7: /cg2_6/prodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description

No matches found					

Search completed: April 20, 2004, 21:13:05
Job time : 19.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: April 20, 2004, 20:45:13 ; Search time 43 Seconds
(without alignments)
11677.204 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689
Sequence: 1 ctcggtgcgcgtccgccac.....actcttaagatcatatcccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%

Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+nzp.model -DEV=xjp
-O=/cg2.1/USPTO.spool.p/US09787016/runat.20042004.154657.16941/app.query.fasta_1.2759
-DB=PIR_78 -OPMT=fastran -SUFFIX=pct.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdd -LIST=1000
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09787016 @CGN 1.1 66 @runat.20042004.154657.16941 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
No matches found					

Search completed: April 20, 2004, 20:57:39
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:41:28 ; Search time 28 Seconds
(without alignments)

9707.365 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689
Sequence: 1 ctcgtrgscgcgcgcacac.....acctctaagatcatatcccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%

Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xip
-O=/cgn2.1/USPTO.spool.p/US09787016/runat.20042004.154656.16910/app.query.fasta_1.2759
-DB=SwissProt.42 -QFMT=fastran -SUPPLY=pct_rap -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN.1.1.28 @runat.20042004.154656.16910 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEIOBURY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
No.						

No matches found						

Search completed: April 20, 2004, 20:51:28
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:44:43 ; Search time 129.5 Seconds
(without alignments)
12718.185 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcgtggtgcgcgcgcgcac.....actctaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xjp
-Q/cgnt2.1/USPTO.spool.p/US09787016/runat.20042004.154657.16922/app.query.fasta_1.2759
-DB=SPREMBL.25 -OPMT=fastan -SUFFIX=pct.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -INITs=bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=1000 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1.1 142 @runat.20042004.154657.16922 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL.25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirts:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query		Match	Length	DB ID	Description
	Score	Query				
-----	-----	-----	-----	-----	-----	-----

No matches found

Search completed: April 20, 2004, 20:56:00
Job time : 129.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:46:13 ; Search time 29.5 Seconds
(without alignments)
9135.172 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689
Sequence: 1 ctcggtgcgcgcgcgcac.....actctaagatcatatccgcg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%
Maximum Match 100%

Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09787016/runat_20042004_154658_16977/app_query.fasta_1.2759
-DB=Issued_Patents_AA -OFMT=fastan -SUFFIX=pct.ra1 -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -INITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pct -NORMext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09787016@cgn2_1_1_29@runat_20042004_154658_16977 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*

1: /cgn2_6/prodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PTUS.COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description
No matches found			

Search completed: April 20, 2004, 20:58:51
Job time : 29.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:56:04 ; Search time 115 Seconds

(without alignments)
12513.164 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689
Sequence: 1 ctcggtgcccgcgcgccac.....actcttaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1124875 seqs, 275673149 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%

Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=fixme+ n2p.model -DEV=x1p
-Q=/cg2_1/USPTO.spool.p/US09787016/runat_20042004_154659_17051/app.query.fasta_1.2759
-DB=Published Applications AA -QFMT=fastan -SUPFIX=pct.rapb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNIT=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human0.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000
-USER=US09787016 @CGN 1.1 17 runat 20042004_154659_17051 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1: /cg2_6/prodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cg2_6/prodata/1/pubppa/PCT_NEW_PUB.pep:*
3: /cg2_6/prodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cg2_6/prodata/1/pubppa/US07_NEW_PUB.pep:*
5: /cg2_6/prodata/1/pubppa/PCTUS_PUBCOMB.pep:*
6: /cg2_6/prodata/1/pubppa/US08_NEW_PUB.pep:*
7: /cg2_6/prodata/1/pubppa/US08_PUBCOMB.pep:*
8: /cg2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
9: /cg2_6/prodata/1/pubppa/US09C_PUBCOMB.pep:*
10: /cg2_6/prodata/1/pubppa/US09C_PUBCOMB.pep:*
11: /cg2_6/prodata/1/pubppa/US09C_PUBCOMB.pep:*
12: /cg2_6/prodata/1/pubppa/US09C_PUBCOMB.pep:*
13: /cg2_6/prodata/1/pubppa/US10A_PUBCOMB.pep:*
14: /cg2_6/prodata/1/pubppa/US10C_PUBCOMB.pep:*
15: /cg2_6/prodata/1/pubppa/US10C_PUBCOMB.pep:*
16: /cg2_6/prodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cg2_6/prodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cg2_6/prodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

No matches found

Search completed: April 20, 2004, 21:17:07
Job time : 116 secs

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
---------------	----------------	-------	--------	----	-------------

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:40:53 ; Search time 119.5 Seconds
(without alignments)
12342.242 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtgcccgcgcgcac.....actcttaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=x1p
-Q=/cgn2.1/USPTO.spool.p/US09787016/unatc.20042004.154656.16900/app.query.fasta_1.2759
-DB=A_Geneseq.29Jan04 -QFMT=faetan -SUFFIX=pct.rag -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09787016@cgn2.1.91@unatc.20042004.154656.16900 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq.29Jan04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

* Result			
No.	Score	Query Match Length DB ID	Description

No matches found

Search completed: April 20, 2004, 20:50:20
Job time : 119.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:12:19 ; Search time 45 seconds
(without alignment)
4305.071 Million cell updates/sec

Title: US-09-787-016a-4
Perfect score: 3228
Sequence: 1 MDCKHLSNEAPRAIKFTS.....LKSCVGLMLALSYSPRPW 614

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_podent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	323.5	10.0	2016	09VG78	09VG78 drosophila
2	228	7.1	2649	09W072	09W072 drosophila
3	228	7.1	2669	09W071	09W071 drosophila
4	225	7.0	2669	09SVB8	09SVB8 drosophila
5	210.5	6.5	5442	0818F0	0818F0 euploies cr
6	209	6.5	663	09W352	09W352 drosophila
7	200.5	6.2	3146	09VUB5	09VUB5 drosophila
8	200	6.2	366	08T3Y1	08T3Y1 drosophila
9	198.5	6.1	424	074508	074508 schizosacch
10	196.5	6.1	645	08VDN7	08VDN7 mus musculu
11	191.5	5.9	2492	07YQM3	07YQM3 pongo pygma
12	188.5	5.8	569	09G6T5	09G6T5 homo sapien
13	188.5	5.8	670	09BM08	09BM08 homo sapien
14	188.5	5.8	671	0724V5	0724V5 homo sapien
15	188.5	5.8	2764	0727D6	0727D6 homo sapien
16	187.5	5.8	1893	08CJ14	08CJ14 rattus norv

17	187.5	5.8	2781	4	09UG2	09UG2 homo sapien
18	183.5	5.7	2492	6	07YQM4	07YQM4 pan troglod
19	179	5.5	385	5	017909	017909 caenorhabdi
20	179	5.5	702	4	092541	092541 homo sapien
21	178.5	5.5	451	5	09NLC1	09NLC1 caenorhabdi
22	177	5.5	443	11	08C969	08C969 mus musculu
23	177	5.5	473	11	08C980	08C980 mus musculu
24	177	5.5	1110	13	091255	091255 petromyzon
25	170.5	5.3	405	5	045410	045410 caenorhabdi
26	170.5	5.3	878	13	08AW4	08AW4 xenopus lae
27	169.5	5.3	781	4	08NAP7	08NAP7 homo sapien
28	169.5	5.3	17352	5	095YM2	095YM2 procambaris
29	169	5.2	5412	5	09W596	09W596 drosophila
30	168	5.2	3279	5	09N4B9	09N4B9 caenorhabdi
31	167	5.2	5327	5	076891	076891 drosophila
32	166	5.1	688	2	09X4J3	09X4J3 ehrlichia c
33	165.5	5.1	563	13	07SZK6	07SZK6 brachydanto
34	165.5	5.1	563	13	07SZB4	07SZB4 brachydanto
35	165.5	5.1	2400	4	081WP2	081WP2 homo sapien
36	165.5	5.1	2416	4	081WP1	081WP1 homo sapien
37	165.5	5.1	2432	4	081WP0	081WP0 homo sapien
38	165.5	5.1	2448	4	081WN9	081WN9 homo sapien
39	165.5	5.1	2464	4	081WN8	081WN8 homo sapien
40	165.5	5.1	2480	4	081WN7	081WN7 homo sapien
41	165	5.1	572	6	08MKF9	08MKF9 canis fam11
42	165	5.1	4498	13	09J291	09J291 fugu rubrip
43	165	5.1	8081	5	072120	072120 caenorhabdi
44	164.5	5.1	510	5	045407	045407 caenorhabdi
45	164.5	5.1	1444	5	09VTN2	09VTN2 drosophila

ALIGNMENTS

RESULT 1
ID 09VG78 PRELIMINARY; PRT; 2016 AA.
AC 09VG78;
DT 01-MAY-2000 (TRMBLrel. 13, Created)
DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
DB CG6525 protein.
GN SPP OR CG6525.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins K.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri1 J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,
RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Fabros B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downe M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glaser K.,
RA Goleser A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwam C.,
RA Jalili W., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodaira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svyrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AB003695; AAF54807.1.
 DR FlyBase: FBgn0038041; Snp.
 DR GO: GO:0003677; P:DNA binding; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR006576; BRK.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD.1.
 DR SMART: SM00592; BRK.1.
 DR SMART: SM00249; PHD.1.
 DR PROSITE: PS01359; ZF_PHD_1.1.
 DR PROSITE: PS0016; ZF_PHD_2.1.
 SQ SEQUENCE 2016 AA; -221598 MW; FB61A32EA710F5B CRC64;

Query Match 10.0%; Score 323.5; DB 5; Length 2016;
 Best Local Similarity 22.7%; Pred. No. 1.5e-13;
 Matches 160; Conservative 86; Mismatches 253; Indels 205; Gaps 30;

8 SNEBPAKXK---PTSEKFRKTKWGRRTTI---AKRBAQD-TEADPEEQPO----- 53
 Db ADEIPIKLVKLESTADLDPPPSLPNITPAPMGOVGNAADAPNNIMDNEDTIT 698
 639 ADEIPIKLVKLESTADLDPPPSLPNITPAPMGOVGNAADAPNNIMDNEDTIT 698
 54 ---QHNLISLRSGRPKRTREVEEPLTTVRRKKQNVVSLIEDSEPTSTVTDVETAS 109
 Db ADPLQHVVLLEEDKQ-PEAEVVKQVLAET-EPGLDAIVSWPTSIIEVDVPOAHTNLLP 756
 699 ADPLQHVVLLEEDKQ-PEAEVVKQVLAET-EPGLDAIVSWPTSIIEVDVPOAHTNLLP 756
 110 EGSV-ESSSEIRSGV-----SDSLGKHHPASSEKAKGSEBEDTSDSDGLTKELON 163
 Db NASLTERPQMSLTPFACSTPSRSVAASTPPTSAAVVGEGYV-----ITLPPY- 805
 757 NASLTERPQMSLTPFACSTPSRSVAASTPPTSAAVVGEGYV-----ITLPPY- 805
 164 RLRRKREDEPERSLGSQNLRRKREDESAETSGVOIGSABD-----RPLCTQDEPEAS 219
 Db EAPTRAKRRAQFSPAPMAATSSSDAGNLSFGESSLDASINQPL----- 850
 806 EAPTRAKRRAQFSPAPMAATSSSDAGNLSFGESSLDASINQPL----- 850
 220 QGPVSQSEITDDIENQIBKATQNTENPREAGPK-----DEC----- 258
 Db ---NTSSLSN--DSQPSGPRKPRPREPSMARSTAPRSKCLDASQNNDDASES 901
 851 ---NTSSLSN--DSQPSGPRKPRPREPSMARSTAPRSKCLDASQNNDDASES 901
 259 ---EYVDPAALVCICQPHNNRPMTCDCGCEPHFDCCGICISARBLERNEDTICNC 316
 Db QEDDDDPKMLKICICQPHNNRPMTCDCGCEPHFDCCGICISARBLERNEDTICNC 316
 902 QEDDDDPKMLKICICQPHNNRPMTCDCGCEPHFDCCGICISARBLERNEDTICNC 316
 317 TILQVQDRTNSATWQSGSGSVGADGTCTSIGTVQSKSGEDGI--KRIERKAPMS 374
 Db VKRQ-----EERSQPRITMLVTRPT--QPRDRPSETVLTTLTAIVQVAPAS 1008
 962 VKRQ-----EERSQPRITMLVTRPT--QPRDRPSETVLTTLTAIVQVAPAS 1008
 375 G-KKKLKIQPVVEAP-----GAPK-----CTGPG----- 398
 Db APRRTLPVLTIVASSPMRIIPAKPAKKEPTGAIHQOQOQOQINFIRIGSPGKRISETLCV 1066
 1009 APRRTLPVLTIVASSPMRIIPAKPAKKEPTGAIHQOQOQOQINFIRIGSPGKRISETLCV 1066
 399 -CSSVAGQDQSVYCSNDCTIKAAATMRFLSSGK-----EOKTPKPK 439
 Db VCKRPASTSSVYCGECIRKTAQSAIQAAATKGLPQNAQASLIANSFPAKRRKKKOL 1128
 1069 VCKRPASTSSVYCGECIRKTAQSAIQAAATKGLPQNAQASLIANSFPAKRRKKKOL 1128
 440 VK-----TKPEK---FSLPKCSVOVGIKISSVHKRLASEKRENPVKYMLASRSBT 487
 Db FEDVLRQADTYISKVERINIVFERKSGRVTITGHMAFAH-QFRKMLQENISFEVLPSGTVOG 1187
 1129 FEDVLRQADTYISKVERINIVFERKSGRVTITGHMAFAH-QFRKMLQENISFEVLPSGTVOG 1187

Qy 488 SGRNACSSSTPSPASDHNVA-----KPEKPKPTALSPTLSTKTYHPKGFPGSPH 542
 Db 1188 ADAPKRLIKAPBARATTSBPATVGAQKP--PGRPAKS-----HQNTTVQASH 1236
 Qy 543 HLGSCGLSRTTRVGLVLTIVASSSLPARSRVQDASGPOVPLPS 586
 Db 1237 QLG-----ISSVRLPAKKDKKXTPTVQAPT 1262

RESULT 2

ID Q9W0T2 PRELIMINARY; PRT, 2649 AA.
 AC Q9W0T2.
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG32346 protein.
 OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Adamatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley B.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Butris K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.B., Gary N.S., Galbraith W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.R., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svyrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RC Celnik S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Adamatides P.G., Brandon R.C., Rogers Y.,
 RA Banton J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Dou L.B., Doyle C., Dresnek D., Fartan D.,

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
RA Paciel J., Parasas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RU Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker S.B., Prochuk S.B., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield B.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RU Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.B., Gibbs R.A., Rubin G.M., Venter J.C.;
RU Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RU Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
RN EMBL: AB003467; AAF47361.2; -
RA FlyBase: FBgn0000541; E(bx).
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; P:dna binding; IEA.
DR GO: GO:0005489; P:electron transporter activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000637; AT hook.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR000345; Cytochrome b5.
DR InterPro: IPR004022; DDT-dom.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR001965; ZnF-PHD.
DR Pfam: PF02178; AT hook; 1.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02791; DDT; 1.
DR Pfam: PF00628; PHD; 3.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00249; PHD; 3.
DR SMART: SM00249; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS00190; BROMODOMAIN_2; 1.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01359; ZF-PHD_1; 1.
DR PROSITE: PS00016; ZF-PHD_2; 2.
DR PROSITE: PS00016; ZF-PHD_2; 2.
SQ SEQUENCE 2649 AA; 298507 MW; C0D7F7E015EA5403 CRC64;

QY 220 OGPNVQSESTD-DIENQLEKATQNTENPRE-AGKPKP----- 256
DB 2377 ASPDSQSNERSGEPNLPFKRTE---VONPRHAGAPKPLTKKKEKYCICRTPYDDTRF 2433
QY 257 -----ECG-VYDPNALYCICRQPHN-NRPNICC 282
DB 2434 YVGCDCLSNFWPHGDCVSTTBASKSLSPFCICDCKRARETQOLYCSCRQPNBSOFTYCC 2493
QY 2493 DRCEMFPHGDDCVGISBARGRLLEKNGEDYICPNTTILQVDETFNSATN 331
DB 2494 DRCQDPFRCVCGILQSRAPFL-----DEVYCEC-----QRNDNMAAN 2533
RESULT 3
OGPW071 PRELIMINARY; PRT: 2669 AA.
ID OG0701
AC OG0701
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DS CG32346-PA.
DS CG32346-PA.
GN E(BX) OR CG32346.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.C., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Chang M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Plamkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borzova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mouton G., Milshina N.V., Modyarty C., Morris J., Mostrefi A.,
RA Mount S.M., Moy M., Murphy B., Modyarty C., Morris J., Mostrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paciel J.M.,
RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spiraling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodruff M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RU Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker S.B., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celinker S.E.,

RA Clamp M.B., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA Harris N.L., Krommell B., Marshall B., Milburn G.H., Richter J.,
RA Russo S., Searle S.M.J., Smith E., Shu S., Smlnick P.,
RA Whitefield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA Lewis S.B.;
RT "Annotation of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO03467; AAN11431.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR004022; DPT_dom.
DR InterPro; IPR006209; EGF_1ike.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; I.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DPT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM0249; PHD; 3.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
DR PROSITE; PS50016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300684 MW; 261FF7B7CEBF781B CRC64;
Query Match 7.1%; Score 228; DB 5; Length 2669;
Best local similarity 22.1%; Pred. No. 9.1e-07;
Matches 77; Conservative 52; Mismatches 98; Indels 122; Gaps 14;
QY 93 DSSEPTSTVTDV-ETASGVSSESSSSTRSGVSDSL-----GKEHPASSEKAKGAE 145
DB 2217 EHDEPTNLAGDIDSETDLENKQNESFVVTGVIQKISNALKQGNLSPLEBKJVCWQKQ 2276
QY 146 EDTSDSDSGLTJLKE--LQNRLLRKRRQBPVERSLRSGQ-----NRLRKRR 190
DB 2277 QENANSTNEMETCSRGSVNEBALTPSRQTDTEWKIRTSLLRPNMTTSSQPNRLKQNR 2336
QY 191 --BEDSAFTGVSVOIGSAEDRPLCKQ-----EPAS 219
DB 2337 SKNDEVALGROKQSLERHKLKKNLLRKSLLERLVQSEIHEDVTKVQBRVRLPSN 2396
QY 220 QGPVQSSTFD-DIENQLEGKATQGTENPRE-AGKPP----- 256
DB 2397 ASPDQSENERSGERNLDFKYTE--VQNPRIHAGRPFKLLTRKKEKLYCICRTPYDDTKF 2453
QY 257 -----ECG-VYDPNALYICICROPHN-NRFMLCC 282
DB 2454 YVGCDDLGNMFGDCVSTTEBASKKLSFICIDCKRARETOOLYSCQPIDESGFYICC 2513
QY 283 DRCEWFGDCVGISEANGRLLENGEDYICPNTTILQVODETNSATN 331
DB 2514 DKQDMWFGRCVGIQSLSEAFI-----DYVCEEC-----QKKNDAAN 2553
RESULT 4
Q95VB8

ID Q95VB8 PRELIMINARY; PRT; 2669 AA.
AC Q95VB8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Nucleosome remodeling factor large subunit NURF301.
GN E(BX) OR NURF301 OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR CG32478.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21468389; PubMed=11583616;
RA Xiao H., Sandilczopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
Fu D., Wu C.;
RT "Dual functions of largest nurf subunit nurf301 in nucleosome sliding
and transcription factor interactions";
RL Mol. Cell 8:531-543(2001).
DR EMBL; AF417921; AAL16644.1; -
DR FlyBase; FBgn000541; E(bx).
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR004022; DPT_dom.
DR InterPro; IPR006209; EGF_1ike.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; I.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DPT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
DR PROSITE; PS50016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300687 MW; 6B4925AFFF489D6F CRC64;
Query Match 7.0%; Score 225; DB 5; Length 2669;
Best local similarity 22.3%; Pred. No. 1.5e-06;
Matches 78; Conservative 53; Mismatches 96; Indels 122; Gaps 15;
QY 93 DSSEPTSTVTDV-ETASGVSSESSSSTRSGVSDSL-----GKEHPASSEKAKGAE 145
DB 2217 EHDEPTNLAGDIDSETDLENKQNESFVVTGVIQKISNALKQGNLSPLEBKJVCWQKQ 2276
QY 146 EDTSDSDSGLTJLKE--LQNRLLRKRRQBPVERSLRSGQ-----NRLRKRR 190
DB 2277 QENANSTNEMETCSRGSVNEBALTPSRQTDTEWKIRTSLLRPNMTTSSQPNRLKQNR 2336
QY 191 --BEDSAFTGVSVOIGSAEDRPLCKQ-----EPAS 219
DB 2337 SKNDEVALGROKQSLERHKLKKNLLRKSLLERLVQSEIHEDVTKVQBRVRLPSN 2396
QY 220 QGPVQSSTFD-DIENQLEGKATQGTENPRE-AGKPP----- 256
DB 2397 ASPDQSENERSGERNLDFKYTE--VQNPRIHAGRPFKLLTRKKEKLYCICRTPYDDTKF 2453
QY 257 -----ECG-VYDPNALYICICROPHN-NRFMLCC 282
DB 2454 YVGCDDLGNMFGDCVSTTEBASKKLSFICIDCKRARETOOLYSCQPIDESGFYICC 2513

146 EDTSDSDSLTKE-----LQNLRRKREQPVERSLRGSONLRKREEDSAETGS 199
 719 -----LYIEVVRVPVLTQDLRLQGLHAIWQDHTVASQ-----QQQQQPGQAGGT 764
 200 VOIGSAGDRPLCKQEPF-----ASQ-----PVSOSET-DDIE 232
 765 TNPQAQO-----VQPPQOWSGIGITVSGSQGPPTAVGYSYFGQOINASQADDDH 819
 233 NQLSGKATQO-----NRENPREAGKPKPECEVNDMLYCIQRPNNRMICCDRECEMP 290
 820 SAISSSSMGLASTDIDPEGETEPAPEARAEEDSVTRICELTHDDGTICCDKCSAMQH 879
 291 GDCVGISARGLRLERNEDYICPNCTILQVDETNCSATNEOD-----SGC 337
 880 VDCMGIDR-----QNIPEBTWCELCOPRAV-DKARALQKQKKEHLVATQAANA 932
 338 RSVGADGDTCTSIGTVEKXSGEDQ-----GIKRI-----EKAANDSGKKK 378
 933 AAVAAGTTLGSLGSLPMSBELQRLASGLNGFPATGTGMSKKKTKTENSGSTTLKK 992
 379 LKITQPVREAPGAPRCIGRCSSVAPGPSVYCSNCLKHAATMRPLSSGKEQTKPKPE 438
 993 TK--KSAVGMGEKKNASGSGTPTG-----SSGRTSKKSKR 1026
 439 KVKT 442
 1027 KSKS 1030
 RESULT 8
 08T3Y1 PRELIMINARY: PRT: 366 AA.
 AC 08T3Y1: 09W533:
 DT 01-JUN-2002 (TREMBlrel). 21, Created)
 DT 01-JUN-2002 (TREMBlrel). 21, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel). 25, Last annotation update)
 DB AT26187P (CG17440 protein).
 DB CG17440.
 OS Drosophila melanogaster (Fruit fly).
 CC Buxarotia, Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese E.,
 RA George R., Gonzalez M., Guerin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.B., Rubin G.M.,
 RA Ceiniker S.,
 RA Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.P., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Baer A., Bendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butte K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Doonan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegwam C.,
 RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kension J.A., Ketchum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh N.V., Mobarry C., Moritz J., Mostrelti A.,
 RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Mostrelti A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamas I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Friese E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Idegwam C., Jaisli M., Kruse D., Li P., Mattei B., Mostrelti A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochick S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.B.,
 RA Clapp M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seattle S.M.J., Smith B., Shu S., Smutniak F., Whitfield B.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RN Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY069443; AAL90181.1; -
 DR EMBL: AB003446; AAF46482.2; -
 DR FlyBase: FBgn0030120; CG17440.
 DR GO: GO:0003677; P:DNA binding; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 1.
 DR SMART: SM00249; PHD; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS50016; ZF_PHD_2; 1.
 SQ SEQUENCE 366 AA; 42048 MW; CB83AAFEAC514CF3 CRC64;

Query Match 6.2%; Score 200; DB 5; Length 366;
 Best Local Similarity 28.3%; Pred. No. 6; 7e-06;
 Matches 54; Conservative 22; Mismatches 65; Indels 50; Gaps 7;

```

QY 267 YCICRQPHNNRPMICCCRCERMFHGDVCGISBARGLLERNGEDYICPNC-----TI 318
DB 40 YCICRSSDSCSFPMIGCCGCEEMHGDICIEITEKDAEHI-----KNYYCRRCKKEMPELOTI 95
QY 319 LQVODETNGSATNEOD-----SGCRSVAGDGTCT-----SIGTVEQSGED 360
DB 96 FRUVATERBAASNAASSTSLNAPGVPSGAAPAAAPVAATITSSQAAPPTTAARAKKNSA 155
QY 361 QGIGRIEIKANPSGKKLK-----IFQPV-----EAPGAPKICGPGCCSSVA 403
DB 156 QEPK-----ESQPTQAGTKKDKAAPTNSVQSPRAVSPFIFLNPLOIGIQOCHGPMCCSHA 212
QY 404 QPDVSYCCND 414
DB 213 RPSKRYCSDEC 223

RESULT 9
Q74508 PRELIMINARY; PRT; 424 AA.
AC 074508;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE SPCC594.05C.
GN SPCC594.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL013523; CAA20664.1; -.
DR PIR; T41449; T41449.
DR GeneDB; Spombe; SPCC594.05C; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS0016; ZF_PHD_2; 1.
SQ SEQUENCE 424 AA; 48683 MW; 2963605C3DFCC0B9 CRC64;

Query Match 6.1%; Score 198.5; DB 3; Length 424;
Best Local Similarity 24.1%; Pred. No. 1e-05;
Matches 95; Conservative 46; Mismatches 139; Indels 115; Gaps 18;

QY 89 VSLSDSSPTSTVLD--VETASGVSSESSIRSGPVDSIGKHPASSERAKGGEERH 146
DB 20 VKPERSNR---CTIDFHITANNEEKDANI-----LNKSVKQVEEVNG---HV 65
QY 147 DTSDDSDGLTKLKIQLNRRKREGEVYERSLRSGONLRKRRREDSATGSSVOIGSAB 206
DB 66 DSSSTETD-IEMQVIQOPTIPCK--PVSAAHRRGP---KKRGGANS---QLNLSTAD 114
QY 207 QDRPLCKBPESASQPVQSETDDIENQLEGKATQNTENRERAGKPRDEKVVDPNAL 266
DB 115 HGRP-----L 119
QY 267 YCICRQPHNNRPMICCCRCERMFHGDVCGISBARGLLERNGEDYICPNCITLQVODITN 326
DB 120 YCICRQPHNNRPMICCCRCERMFHGDVCGISBARGLLERNGEDYICPNCITLQVODITN 326
QY 327 GSATNEQSGCTSVAGDGTCTSIGTVEQSGED---QGIGRIEIKANPSGKKLKIFQ 383
DB 171 GITTAKR--CGLRECSNPTRENSVCSDKHGVDFREKVK---LSTVEPSATKQLVLPFA 225
QY 384 PVEARGAPKICGPGCCSSVAQPDVSYCCNDICLKHAAATMRPLSGKGGKTPKKEKVKTK 443

```

```

DB 226 KKRSEFQULGVGVPRPSQVPEPVY-----NFEIEBANRLANIEIQLNKEKVAASNK 278
QY 444 PEKSLPRCSVQVGIKISSVHKRLASERENPVKK 478
DB 279 -----KIPQL-IKSSRRRAVLAKEREG-IKK 304

RESULT 10
Q8VDN7 PRELIMINARY; PRT; 645 AA.
AC Q8VDN7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN PALZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RA Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC021489; AAH21489.1; -.
DR MGI; MGI:244408; Palz.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR006209; BGF_1like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR PROSITE; PS00633; BROMODOMAIN 2; 1.
DR PROSITE; PS01186; BGF_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS0016; ZF_PHD_2; 1.
DR Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 645 AA; 72167 MW; 6486F34F339820F9 CRC64;

Query Match 6.1%; Score 196.5; DB 11; Length 645;
Best Local Similarity 20.1%; Pred. No. 2.4e-05;
Matches 76; Conservative 64; Mismatches 125; Indels 113; Gaps 13;

QY 43 TRADSEQQPOQHNLSTR-----SGROPKTERVEFLTYRRRGKKVVPVSL-EDS 94
DB 146 TVLSSECPQPOPVAVIQLQOVVLSQISQVVAQIAQSGVGVQIKLQVPIQVQNS 205
QY 95 SEPTSTV-----DVETASGVSSESSIRSGPV----- 124
DB 206 AAQTSVVTVQAASVQBLQRYQLRQQQKQKQIIFERHETTLAASNSGSEIILQGVCHK 265
QY 125 -----SDSLGKHPPASSERAKGGEER-EDDT 148
DB 266 HNAVTEHLKQKMTTPAREBENQRMVGNQVKKYILDKIKERQDAKKRKEEVEYQGR 325
QY 149 SPSDDGILT-----KSLQNRLLRKR-----QEPYERSLRSGONLRKRRER 192
DB 326 SKQNSAKSLALLFKRKEQLKALRLDLKELOIQVQBSLKRLL-----MKKERENA 380
QY 193 DSARTGVSQIGSABQDRPLCKQBPESASQPVQSETDDIENQLEGKAT-----QNTG 245
DB 381 QAVQANAAASVTPSPAPVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 440
QY 246 ENPREAGKPRKPC-----EYVDNALYICRQPHN-NRPMICCCRCERMFHGDVCGISB 298
DB 441 EERKSKSKKKMMISTSKAKQOTRLYICITPYDSKFFYIGCDRCQNMVHGRGCVGIIQ 500

```

Qy 299 ARGILLERNGEDYICPMC 316
Db 501 SEADLI----DEVCPQC 514

RESULT 11
ID 07YOM3 PRELIMINARY; PRT; 2492 AA.

AC 07YOM3; (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DB ATRX protein.
OS ATRX.
OC Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pongo.
OX NCBI_TaxID=9600;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22763540; PubMed=1277533;
RA Kltano T., Schwarz C., Nickel B., Paabo S.;
RT "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and
RT Chimpanzees";
RL Mol. Biol. Evol. 20:1281-1289(2003).
DR EMBL; AB102643; BAC01112.1;
SQ SEQUENCE 2492 AA; 282614 MW; 48EC97106D0C9P20 CRC64;

Query Match 5.9%; Score 191.5; DB 6; Length 2492;
Best Local Similarity 18.9%; Pred. No. 0.0028;
Matches 108; Conservative 114; Mismatches 199; Indels 151; Gaps 24;

Qy 3 DKHLNBARAKATPTSKERKRTTAKREAGADTADSDSQPOQHNLSTARS 62
Db 972 DQDETSEDDKQSKGTEBEKRTSDFKKCVI-KNQYVSSSDTEKLPREKIC--- 1026
Qy 63 GROPKTRVEFLTVRRRGKQNVSLDESEPTSTVTDETASBG---SVESSEH 119
Db 1027 -HPEGIQIQKIGTTGDKRKKK-----IRDTSKKKKSLYAKSTGKSCSSBK 1080
Qy 120 RSGPVSDSLGKHPASSBRKAKGEBEEDTSDSDGLTLK-----LQNLRR----- 167
Db 1081 KS--KNGAVGKRRKCTLLGKSSRRQDCSSDTEKYSKEDGCSNDRKLRILERR 1138
Qy 168 ----KREGEPEVRSLRGS----QNLKRRKREEDNAETGSGVIGAEQDRPLCKQEPAS 219
Db 1139 NLSSGRNTKEIQSGSSSDAEBSSEDNKKKQRTSSKKKAVTV--KXKGRSLRTSTK 1196
Qy 220 QGPVQSFTDIDENQLEGATQNTENPRBAKPECEVYDPNALYCICRQPHNNRM 279
Db 1197 QADITSSSSSDIEDDQNSIGGSSD--QIKYVTEMLVLSHRTG----- 1242
Qy 280 ICDCRCBEMFHGDCVGISBARGRL-----ERNGEDYICPNCTILQVQ---DETN 326
Db 1243 -----QSSGD-----EALSKSVPTVDDDDNDPENRIAKWMLERIKANLSDRD 1289
Qy 327 GSATWEGSGCRSVAGDTGCTSIGTVQKSGEDGIGRIKIRKANPSGKKLKIPOPV 386
Db 1290 GSSDDEPEEGKRTKON-----EENPDDEBA-KNQVNSESDSDSESKK----- 1333
Qy 387 BAPGAPKICIGPCSSVAQPDVYCSNDCTIKHAAATMRPLSGKEQKTPKKE--KVKTYP 444
Db 1334 -----PR-----YRHLRLH-KLTVSNGSGEKRTKPKHKKVKGKN 1370
Qy 445 EKPLPKCSVOYGIKISSVHKRLASBKQENPVKVMLASRSR-----TSGEACESSTP 499
Db 1371 RR-----KVSEDSR--DSDPQSGVBSVESESDQRPRTSAKKALEENQR 1417
Qy 500 SW-----ASDHYNNAVKPEKPEK 517
Db 1418 SYKOKKKRRRIKQVDBSSSENKSNSEBEEK 1449

RESULT 12
ID 096G15 PRELIMINARY; PRT; 569 AA.

AC 096G15; (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DB Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009449; AA09449.1;
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 569 AA; 62857 MW; 253856PC544PDBEA CRC64;

Query Match 5.8%; Score 188.5; DB 4; Length 569;
Best Local Similarity 21.6%; Pred. No. 7.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

Qy 28 GFRRTT-----IAKREAGADTAD-----PSEQPOQHNLSTARSGRQPKTERVEBFL 76
Db 52 GIKRRTPLAKNSVSRKARKASDLDQASVPSBENSSSESEKTSDDPTPEK--KAA 109
Qy 77 TVRRRG-----KNQVPSLDESEPTSTVTDETASBGVSSSRIRSGPVSDSLK 130
Db 110 VAPAPRGPLGKRRKAPASDSDKADSDAKPEPVAAASASSSSSSSSSDSVSYK 169
Qy 131 EHPASS-----KAKGEEED-----TSDSDGL-TLKELONR-----LKR 167
Db 170 KPPRRKPEKPLPRGRKPPRPPSSSSSDSDSDVDRISEWRDRDEARRRLEBAR 229
Qy 168 KROEPVRSLSGQNLKRRREDSATGSGVOISA-----EODRPLCKQEPK 218
Db 230 RROEHELRRLBOEKERKRRER--ADRGAEKSGSSGDELREDEBPVKKRGRK 287
Qy 219 SGPVQSFTDIDENQLEGATQNTENPRBAK--PPECEVYDPNALYCICRQPHNN 276
Db 288 GCGPSSSDSER-EAELR-----REAKKSAKPPQSSSTP-----ARPD--- 325
Qy 277 RFLICDCRCBEMFHGDCVGISBARGRLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
Db 326 -----GQKEKRV-----PEKQO 339
Qy 337 GSVAGADGCTCTSIGTVQKSGEDGIGRIKIRKANPSGKKLK-----IFQPVVEAR 390
Db 340 APVAVERT-----RKNSGFSMDKVKKEKPSVEKLOKLSHSEIKALAKVSDP 390
Qy 391 APKICIGPCSSVAQPDVYCSNDCTILK--AAATMRPLSSGKEQKTPKPKVTKPKES 448
Db 391 VRCIL-----NALEHGTQVTSQILQKNTDVATLK-----KIRRKANKDVMKAAEY 441
Qy 449 LPKGVOVGIKISSVHK-RLASBKQENPVKVMLASRSRSGKEAACSSTPSWAD 504
Db 442 TLKRSRVLPKTEAVQVKAKGKREKAEKL---AGBELAGEBAPOEFAEDKVPSTD 495

RESULT 13
ID 09BW08 PRELIMINARY; PRT; 670 AA.
AC 09BW08; (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DB Similar to hepatoma-derived growth factor, related protein 2.
OS Homo sapiens (Human).

OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC000755; AAH00755.1;
 DR InterPro; IPR000313; PWM domain.
 DR Pfam; PF00855; PWM; 1.
 DR SMART; SM00293; PWM; 1.
 DR PROSITE; PSS0812; PWM; 1.
 SQ SEQUENCE 670 AA; 74229 MW; BAA2429146BDBAD CRC64;

Query Match 5.8%; Score 188.5; DB 4; Length 670;
 Best Local Similarity 21.6%; Pred. No. 9e-05;
 Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

OY 28 GFRRTT-----IAKREGAGDTGAD-----PSEOOPOOHINSLRSGRQPKRTREVEFL 76
 DB 154 GLKRTKATLAKMSVSKRARKASDDLQASVSPSEBENSSSESEKTSDDQDTTPK--KAA 211
 OY 77 TTVRRG-----KQNPVSLDESSBPTSTVTDTETASBGSVSSSEIRSGPVSDSLGK 130
 DB 212 VRARRRPLGRKKKKKAPASDSKADSDGAKPEPVAMARSASSSSSSSSSDSVSVK 271
 OY 131 EHPASS-----KANGEBEED-----TSDSDSL-TLKLQNR-----LRR 167
 DB 272 KPFGKRPKAEKPLPKPRGKPKPRPPSSSSSDSDSDDEVDRISWKRRDARRELEARR 331
 OY 168 KREGEPEVRLSGONLRKKRREEDSAFTGSGA-----EODRPLCKQBPAA- 218
 DB 332 RREGEELRLRLEQEKKEKERRER--ADRGARSGSGGSDRLREDEPVKRGKGR 389
 OY 219 SOGVSQSETDIDINOLEGKATQNTBENPREAG--PKPECEYVDPAALYCIGOPHNN 276
 DB 390 GRGPPSSDSEF-BAELE-----REAKSAKKPSSSTP-----ARKP--- 427
 OY 277 RFMIDCDRCBEMFHGDCVIGISEARGLLERNGEDYICPNCTIILQVDETNGSATNEODSG 336
 DB 428 -----GQEKERVR-----PEEKQ 441
 OY 337 CRVAGAGTCTGTSIGTYEOKSGEDGIGRIEKAANSGKKLK-----IFQVPEARG 390
 DB 442 AKPVKVERT-----RKSGEFSMDRKVEKKKESVEKKLQKLHSEIKFALKVSDPD 492
 OY 391 APKCIIGCGSSVAQPDVVCNDILKH--AAATMFLSSGKEOKTKPKKVKTKPEKFS 448
 DB 493 VKRCL-----NALEELGTLYTSQILQNTDVAATLK-----KIRRYKANKOVMEKAAYV 543
 OY 449 LPKCSVOVGIKISSVHR-RLASERENPVKKVMLASSETSGKAACSSSTPSASD 504
 DB 544 TRLKSRVLTGPRLEAVQVKNKAGKEKAEKEXL--AGBELAGBAPQEKADKDPSTD 597

RESULT 14
 Q724V5 PRELIMINARY; PRT; 671 AA.
 AC Q724V5;
 DT 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hepatoma-derived growth factor 2.
 GN HDGF2.
 OS Homo sapiens (Human).
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yu L.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF294267; AAP97281.1;
 SQ SEQUENCE 671 AA; 74316 MW; 66CA046B0E802741 CRC64;

Query Match 5.8%; Score 188.5; DB 4; Length 671;
 Best Local Similarity 21.6%; Pred. No. 9e-05;
 Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

OY 28 GFRRTT-----IAKREGAGDTGAD-----PSEOOPOOHINSLRSGRQPKRTREVEFL 76
 DB 154 GLKRTKATLAKMSVSKRARKASDDLQASVSPSEBENSSSESEKTSDDQDTTPK--KAA 211
 OY 77 TTVRRG-----KQNPVSLDESSBPTSTVTDTETASBGSVSSSEIRSGPVSDSLGK 130
 DB 212 VRARRRPLGRKKKKKAPASDSKADSDGAKPEPVAMARSASSSSSSSSSDSVSVK 271
 OY 131 EHPASS-----KANGEBEED-----TSDSDSL-TLKLQNR-----LRR 167
 DB 272 KPFGKRPKAEKPLPKPRGKPKPRPPSSSSSDSDSDDEVDRISWKRRDARRELEARR 331
 OY 168 KREGEPEVRLSGONLRKKRREEDSAFTGSGA-----EODRPLCKQBPAA- 218
 DB 332 RREGEELRLRLEQEKKEKERRER--ADRGARSGSGGSDRLREDEPVKRGKGR 389
 OY 219 SOGVSQSETDIDINOLEGKATQNTBENPREAG--PKPECEYVDPAALYCIGOPHNN 276
 DB 390 GRGPPSSDSEF-BAELE-----REAKSAKKPSSSTP-----ARKP--- 427
 OY 277 RFMIDCDRCBEMFHGDCVIGISEARGLLERNGEDYICPNCTIILQVDETNGSATNEODSG 336
 DB 428 -----GQEKERVR-----PEEKQ 441
 OY 337 CRVAGAGTCTGTSIGTYEOKSGEDGIGRIEKAANSGKKLK-----IFQVPEARG 390
 DB 442 AKPVKVERT-----RKSGEFSMDRKVEKKKESVEKKLQKLHSEIKFALKVSDPD 492
 OY 391 APKCIIGCGSSVAQPDVVCNDILKH--AAATMFLSSGKEOKTKPKKVKTKPEKFS 448
 DB 493 VKRCL-----NALEELGTLYTSQILQNTDVAATLK-----KIRRYKANKOVMEKAAYV 543
 OY 449 LPKCSVOVGIKISSVHR-RLASERENPVKKVMLASSETSGKAACSSSTPSASD 504
 DB 544 TRLKSRVLTGPRLEAVQVKNKAGKEKAEKEXL--AGBELAGBAPQEKADKDPSTD 597

RESULT 15
 Q727D6 PRELIMINARY; PRT; 2764 AA.
 AC Q727D6;
 DT 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Bromodomain PHD finger transcription factor.
 OS Homo sapiens (Human).
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Barak O., Lazzaro W., Lane W., Speicher D., Picketts D.,
 RA Shiekhatcar R.;
 RT "Isolation of human NURF";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY282495; AAP22284.1;
 SQ SEQUENCE 2764 AA; 308079 MW; 9785D7097C2099F9 CRC64;

Query Match 5.8%; Score 188.5; DB 4; Length 2764;
 Best Local Similarity 19.3%; Pred. No. 0.00052;
 Matches 70; Conservative 67; Mismatches 111; Indels 95; Gaps 13;

OY 31 RTTYAKRGAGDTGADSBQPOOHINSLR-----SEOPKRTREVEFLTTVRRGK 84
 DB 2289 QTVSSHV-----PSEAOPTHAQSSKPVQAQSQPSNVQGSFVAVQSPSQ--TRIRSTP 2343

```
QY      85 KNPVSLSDSSEPTSTVTVDVETASBGSVSSSEIRSGP----- 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2344 SLSFGQSQVQTTSQPIPIQPHSLQIPSGQPQSQPVVMKGNNAVIBHLKQKSWTP 2403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 -----VSPSLGKEHPASSEKAGGE-REDTSDSDGLTL----- 158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2404 AERBENQMTVCNQVMKYILDKIDBEKQAKKRRRESEVQKRSKQATYLSALLFNHK 2463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 KELONRLRRKRE-----QEPVERSLRSONRLKRRREDSAEFGSVQIGSAEQD 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2464 EQLAABILKRALADKDLQIEVQZELKRDL-----KIKERDLMQLAQATAV----AAPC 2514
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 RPLCKQZPEASQGFVSQSETDIEN-----OLEKATQGNTEBNPREAGKPE---- 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2515 PPTVPAPAPAPAPPPPPPPAVQHTGLSTPTLPAAQKRRREBKDSSSKSKKMMIS 2574
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 ---CEVTDPNALYICICROPHN-ANPMICCDRCCEWYHGDVIGISEARGLLERNGEDYIC 313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2575 TTSKETTKDVTLYCICIKTPYDESKFYIGCDRCQWYHGRVCVGIQSEAEI----DEYVC 2630
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 PNC 316
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2631 PQC 2633
```

Search completed: April 20, 2004, 21:19:50
Job time : 48 secs

SUMMARIES

ALIGNMENTS

```

30   154.5    4.8   1151     2   T24541      hypothetical pr
31   154.5    4.8   1560     2   T30282      calcium-binding p
32   154      4.8   555      2   T26413      hypotetical prote
33   154      4.8   1390     2   T14004      triA protein - bli
34   153.5    4.8   1849     2   C41859      Iga-specific metal
35   153      4.7   1165     2   T16420      hypotetical prote
36   151      4.7   2282     2   T42717      DNA-binding protei
37   150.5    4.7   1077     2   T21800      hypotetical prote
38   150      4.6   1280     2   G96796      hypotetical prote
39   148      4.6   657      2   S10001      mydII protein - m
40   148      4.6   863      2   T49709      related to glucan
41   147.5    4.6   258      2   T06070      B543 protein homol
42   147.5    4.6   258      2   H85462      B543 like protein
43   147.5    4.6   1166     2   H86341      hypotetical prote
44   147      4.6   1507     2   B47328      natural killer cel
45   146.5    4.5   644      2   S55395      neurofilament prot


RESULT 1
T14449
Probable phd finger transcription regulator - fission yeast [Schizosaccharomyces pombe]
C|Species: Schizosaccharomyces pombe
C|Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C|Accession: T14449
R|Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A|Reference number: Z21994
A|Accession: T14449
A|Status: preliminary; translated from GB/EMBL/DDBJ
A|Molecule type: DNA
A|Residues: 1-424 <RT>
A|Cross-references: EMBL:AL031523; PIDN:CAA20664.1; GSPDB:GN00068; SPDB:SPCC594.05c
A|Experimental source: strain 972h-, cosmid c594
C|Genetics:
A|Gene: SPDB:SPCC594.05c
A|Map position: 3

Query Match          6.1%; Score 198.5; DB 2; Length 424;
Best Local Similarity 24.1%; Pred.No. 0.00042;
Matches 95; Conservative 46; Mismatches 139; Indels 115; Gaps 18;

OY VSLDSEPTSTVTVD--VETASGCVSSSERSISGPVSDSLGEKHPASSSKAGGBEEB 146
Db VKFEDSNR---GITTDFAIETANNBERKANVI-----LNKSVMKEVEEVNG--HV 65
147 DTSDSDSGTLTKELQNRRLARKREHQPVERLSRGSONRLRKRRBREDSAFNGVOIGSAE 206
Db DSSSTETD-IHQVATQQOPTIRK--PVVASRRCP----KKHGNAANS-----QLNLSTD 114
207 QDRSLCKREPASQGPSQSSETDDILENQLEKATQTGTENBRPRAKPEPCENVDPNAL 266
Db HQRP-----L 119
OY YCICRPHNNRFPICCDCRCERWTFEGDCYGISBAKRLLERNGBDYICNYCTILLOVDET 326
Db YCICOKPEDDGSMWLGCOCGBDMFHGTCAWTIPESYNDDLVOQ--YFCPKCT-----EBGK 170
327 GSANNBDSGRSVAGADGDTCTSIGTVBKSGED---OGIGRIEKANPGSKKGKLKIQQ 383
Db GITTWKRK--CRLBECNSPTRPNNSNYCSDKRGVDFPRKVK---LSTYEPSAIKRLVLPA 225
384 FVNEAPGAKPICIGGCSSVAQPDVSVCNDCLLKHAATAWRFLSSGKEQKTPEKRVTKX 443
Db KKBEPNGLGVGTLTLSQVPBPVVY-----NPRIEABARLNABIQLNKKEGVASNKX 278
444 PEKTSLPKCIVQGIKISSVHKRIASEREKNPVKK 478
Db -----KIFQL-KDSRRAAVALAYKERREG-IRK 304

```


RESULT 2

A34736
 nestin - rat
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 13-Jul-1990 #sequence_revision 09-Oct-1992 #text_change 05-Nov-1999
 C/Accession: A34736
 R/Endahl, U.; Zimmerman, L.B.; McKay, R.D.G.
 Cell 60, 585-595, 1990
 A/Title: CNS stem cells express a new class of intermediate filament protein.
 A/Reference number: A34736; PMID:90150286; PMID:1689217
 A/Accession: A34736
 A/Status: preliminary
 A/Molecule type: DNA, mRNA
 A/Residues: 1-1805 <LEN>
 A/Cross-references: GB:M34384; NID:9205663; PID:AAA41665.1; PID:9205664
 A/Note: the authors translated the codon CCG for residue 168 as Gly, ATC for residue 169
 C/Keywords: coiled coil

Query Match 6.1%; Score 197.5; DB 2; Length 1805;

Best Local Similarity 21.2%; Pred. No. 0.0023;

Matches 143; Conservative 99; Mismatches 223; Indels 211; Gaps 35;

```

Oy 2 DDKG---HLSEBARKAIPTSKFRKTWGFRTTIARBGADTEADPSBOQPOOHNL 58
Db 486 RDEQIWLVERKADIRKVENSKAKT-----QESGLDTEFQDSGPIQK-T 534
Oy 59 LRSGRQPKTERVEFLTVARRGKNVPSVLSDS-----SEPTSTVDVETASG 114
Db 535 LKALGEPLMSLKIQNYETA-----GKENCNSTEGHGLTLEGPEKQIPLKSK 590
Oy 115 SSSERKSG-PIV-SDSLGKHPASSRK-----AKG-----GEHEEDTSDSDG 155
Db 591 SKTLTENGVPVLSLILGKEDTETEDQELMSPKGTLKRPSLSKESQEVVRPSKE 650
Oy 156 -----LTLKELQN-----RLRKRKEQEP----- 173
Db 651 TAFKESQHPGLGPRGAEQDMLERLVKEDQSPRPSEEDQACRPLQKEMQEPGLY 710
Oy 174 ---VERSL-RGSQNLKRRREEDSAETGVSQISAEQDRPLCKQEPASQGVQSQT 228
Db 711 EGQILERLIRKESQSLRSP--EEDQENG-----RSLQKX-----NQEP 754
Oy 229 DD--IENOLEKATGQ--NTEENPREAGRP-----KPCGEVVDNALYICQPH 279
Db 755 EDQMLERLIRKESQSLRSPENQR--IGKPLERENQKSLRYLENOETFPLES 813
Oy 280 ICCDRCER-----WFGDCVGIS-----EARGRL 303
Db 814 RSLVEEHEQRVYKLEKVSQDSLSLAENQPLRYLEEDCINKSLBEDTKHSLSL 873
Oy 304 LERNEDYICPNCTILQV-----ODBTNGSATNEQDSGCSVAGDGT 346
Db 874 EDKNGDSIIIPQSEFQVSLRPEBEDQRIVNHLEKESQEFRRSSSEERQVNER 933
Oy 347 CTSTGTQKSGEDQIGKIRIRKANPSGK---KKLKIPOVVAERAPKICIGCCSV 402
Db 934 HESLSVSRK---EDQVBSQLEKESQDSKSLSDSQRTFGL-----EKNAS 984
Oy 403 AOPDVVCSNDICILKHAATMFLSSGKEQ--KTKPEKRYTKPEKPSLPKCSGVQ 461
Db 985 AGQDQ---EEOGLBEFTQTLR--AVGNEQMAVSPPEK--DPE---PLG--NDQ 1030
Oy 462 SVHKRLASEREN-----PYKKV-----MLASRSTSGKAACESTPSWASDH 505
Db 1031 EIASRLGKENGQSLVSLKKGIEYVSLTEBIIIPLETAEDLERKRSIDTQEP 1090
Oy 506 NYNAVKPEKPEKPTAL 521
Db 1091 ARETVEPPDEPPGSL 1106

```

RESULT 3

hypothetical protein B0041.7 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 28-Oct-1999
 C/Accession: T34036
 R/Fulton, R.; Mohlmann, P.
 submitted to the EMBL Data Library, April 1997
 A/Description: The sequence of C. elegans cosmid B0041.
 A/Reference number: 221466
 A/Accession: T34036
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1359 <FULL>
 A/Cross-references: EMBL:AF000196; PID:AA024256.1; GSPDB:GN00019; CESP:B0041.7
 A/Experimental source: strain Bristol N2; clone B0041
 A/Genes: CESP:B0041.7
 A/Map position: 1
 A/Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 5.7%; Score 185; DB 2; Length 1359;

Best Local Similarity 21.4%; Pred. No. 0.0086;

Matches 146; Conservative 88; Mismatches 217; Indels 232; Gaps 33;

```

Oy 8 SNEBAPKAIPTSKFRKTWGFRTTIARBGADTEADPSR--QOPOOHNLRLRSG 65
Db 210 SESEDEKVKSKSKSK-----VKKESSEDEAPKPKTKRKRSKTSSESSSS 261
Oy 66 PRTREVEFLTVARRGKNVPSVLSDSSEPTSTVDVETASR---GSV---ESS 117
Db 262 EKSDREERKESPPKPKKK--PLAVKTLSDREERSESEVEVLPQKKRGAATL 319
Oy 118 EIRSPVDSLCKEPPASERAKGEEBEDTSDSDGLTLGELNRLRKRKEQ--EP 176
Db 320 EKQKSEBAPKESKAKKQESSSGDSBSGSLTV---NRKSKKKEKPKKK 375
Oy 177 SLRGSQNLKRRREEDSAETGVSQISAEQDRPLCKQEPASQGVQSQTVDIEM 236
Db 376 GLIMSSYLQKETTIDARAE-----KERKLEKQKERN--GIVLEGBDILT 425
Oy 237 GNATGATTEB--NP-----REAGKPECEGYDNPALYICR--QPNRRPMIC 282
Db 426 GTSSQKSKSVLIDPSSVDESKRP--VEVH--NSLVRLIKPHQAHGIDFMDCA 480
Oy 283 --DRCEMFWHG--DCVGISRARGRL-----EENGED-----YICPNCT 319
Db 481 SLDRLDTBGSGGILAHCHGL--GKTLQVITFTATYLMHEKIGKCKRVLYVPP 536
Oy 320 -----QVODETNGSATNEQD-----SGCRSVAGDGTCTSIG 351
Db 537 NMFKEPQKLVNDEBELDIDVNEHDSYKTIEDRRBALKAMHSKTPVMII 596
Oy 352 TVEQ--KSGEDQIGKIRIRKANPSGKKLKIPOVVAERAPKICIGCCSVAD 409
Db 597 TVEDDPKKKPKRRRLERAKED-----FRKYLQNPQ-----PDWV 634
Oy 410 CSNDICILKHAATMFLSSGKEQTKPKPKRYTKPEKPSLPKCSGVQGIKISS 469
Db 635 CD-----EAHLKDDSLSKCMVLT----- 655
Oy 470 EKRBNPVKKVMLASRSTSGKAACESTPSWASDHNTAVNAPKPEKPTALSP 529
Db 656 -----LTKRRLCTNGTP-----LQNNLM-- 674
Oy 530 TYHPRAGPFGPSHLLGGCLGSRTEVLGVLYLVASSSLPARSRQDASGPQV-- 588
Db 675 --YHCVNVEFKP-----GLLG--TKTEFANRFVMI--NKGRTKDSPL 721
Oy 589 SLSGWFLKSCVQ-----LMLBAI 606
Db 722 HVLVDHLKKCVDRKQDYRVLTBAI 744

```

RESULT 4

[illegible]

```

RESULT 7
151116
NF-180 - sea lamprey
C/Species: Petromyzon marinus (sea lamprey)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: 151116
R/Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A/Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation
A/Reference number: 151116; MUID:95287814; PMID:7770000
A/Accession: 151116
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1110 <JC>
A/Cross-references: EMBL:U19361; NID:G632548; PIDD:AAA60106.1; PID:G632549
A/Superfamily: neurofilament triplet H protein

```

Query Match	5.5%;	Score 177;	DB 2;	Length 110;
Best Local Similarity	18.9%;	Pred. No. 0.019;		
Matches 102;	Conservative 74;	Mismatches 197;	Indels 168;	Gaps 17

Qy	2	DDKXHL-----SNEEPAKAIPTSKPEPKMTGFRRTTIAKREGGADTADSDPOQ-OH	55
Dh	509	EDRRKRCGEABEEBAEVEEKEEAB-----EABVEAB-ABETABAAEHEEBAEG	559
Qy	56	NLSRRSGRQPKTERVEEPLTVARRGKNPV-----SLBDSSEPTSTVTDV	105
Dh	560	EEBAABEGEABABAEVEEBAI EKAAABAVABVEEBAABAEHEEBAEBAEVEATKE	619
Qy	106	ETABEGSVRESSBIRSGPVSDSLGKHPASSSEKACGE---EEPTSDSDDLTXLQ	162
Dh	620	EVEABAEATBEEGBGAABAEBAEBAEVEEVTSKATTOABEVEEBAEAAEA-----EAB	673
Qy	163	NRLARKREOEPEVERSLRSGQNRNLKCRREBBSAFTSGVQISAEQDRPLCQEPETASQGP	222
Dh	674	EEABEEBAEBOVEAB-----SKGHEEDBSKE-----ADAEDE---ABEEVABEB	716
Qy	223	VSQSEETDIIENOLBEKATQAGTNEENPREAGKPKPECEYDPPALYCIKROPHNRMFICC	282
Dh	717	VTKSDABAEABAEBAEBAAK---SEEEABAEABDABEEBAEBAEBAV-----	758
Qy	283	DRCEMFIFGDCVSGABKRGLLERNGBDYICPNCITLQVQBETNSATNEODSGCRSYGA	342
Dh	759	-----BETRAATEEBA-----	770
Qy	343	DGTDCTSIGTVEOKSGBDQIKGRIBKANFSGKKQKLIPOPVEABAPAPKICIGPCSSV	402
Dh	771	-----KEASDDEKPEBEEVYESBAVAPAPKAPAP-----KAAPKAP- -AKV	812
Qy	403	AQPDVYCSNDCILKHAANAATRPLSSGKEOKTKPKPEKYTKVPEKPSLFCGSOVQIGIKSS	462
Dh	813	ESPTS-----EPDEBPAKAVVEKKGRABAKPA-----	841
Qy	463	VHKILASERENPVKQVMLASRSETSGKEAACESSTPSMASDHNNAVYPERKEKPTALS	522
Dh	842	--KPAALAKCKAPVKEKEEPESPTBEEPKTPAALKPAAKPAKPPAPAKYAEKBEKPAK	899

QY	523 P 523
Db	900 P 900

RESULT 8
JW0057

Gravin - human
C.Species: Homo sapiens (man)
C.Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C.Accession: JN0057
R.Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A.Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial cells
A.Reference number: JN0057; PMID:98269042; PMID:9604001
A.Accession: JN0057
A.Molecule type: mRNA
A.Residues: 1-1684 (<SAT>
A.Cross-references: DDBJ:AB003476; NID:92081606; PIDD:BA119927.1; PID:d1020716; PID:g2081606
C.Comment: This protein regulates cell growth.
F:433-439/Region: nuclear location signal
F:522-527/Region: nuclear location signal
F:591-596/Region: nuclear location signal
F:671-676/Region: nuclear location signal

Query Match	5.3%	Score 172.5;	DB 2;	Length 1684;
Best Local Similarity	21.4%	Pred. No. 0.055;		
Matches 135; Conservative	80;	Mismatches 212;	Indels 203;	Gaps 32

```

0Y 10 EEPAPAI-----KPTSK-----FKTW--GFPRT-----32
0Y 42 EETPBIIOIPESBSNLABELTQPTESQANDIGKVKYKVFVGFYTKKDTEKEDTQOLL 101
0Y 33 TAKREGADITVA---DPS-----BOFOQHNLILRSGROPKTERVEEFLTV 79
0Y 102 TVKQEGEBAAGCHDKDPSLAGEBAASKESBKQ-----STENKPEBTLKR 148
0Y 80 RRRGKNVPSLSDSEPTSTTVTDVETASGESSVESSSETRSGVPSLSLK-----130
0Y 149 QSHAETSPRA---ESQAVAECKBECKEKOEKPEKSABSPTSBTSSTFKKFTQ 205
0Y 131 ----EHPASSERAKGSEEDTSDSDGLTLKELQNLARKQEBVVERSLASQNLRL 186
0Y 206 WAGMRKTSFKPK-EDVEASE-----KKGEQF-----E 235
0Y 187 KRRREED-SAEFGSVQIGSABDRPLCKQF-BASQPVQSSEFTDI---ENOLEGKAT 240
0Y 236 KYDTEEDGAAVYASERULTASBOHNF--QEPASAEHPRLSAETAKVELPSREQVSG--S 290
0Y 241 QGNTENFPRBAGPKF-ECEYVDNALYCLCRPHNNRF--ICDCECFHDDCGIS 297
0Y 291 QSPSEB---KPAFLATEVPDEKI-----EVOHEEVAEVAHVSTVEETBEOKTEV 338
0Y 298 EARG-----RLERNG-----DYICPNCTILQVDEFTNGSATVEODSGCRSVAGD 344
0Y 339 ETAGSVPAEELVBMALPQBAEPAKELVTKLCTVSGEDTQO-ADLSPEKVLSPKPEG 397
0Y 345 TDCTSIGTVBOKSGEDQ-GIKGR-IRKANPSCKKLKTIPQVVEAPGAKFCIG-----396
0Y 398 V---VSEVEMLSQERMTVQCSPLKCLFTSTGLTKLSGKKOKKRGCGDESGEHNVP 453
0Y 397 -----PGSSVAQPDVSYCNSDCLKHAATMRPLSGSGEOKTKPREKV-- 441
0Y 445 ADSPDSQEBOKBSSASSPBBEBITC-----LEKGLAVQOOGKAEBGATSDDEKREG 508
0Y 442 ---TKPEKSLPKCSVQVQIGIKLSVHKRLASERENPVKMLASHSETSGKEACESS 497
0Y 509 VTPWASFKQWVTPKKRV-----RRPSEDEKDELKVSATLSST-----ES 550
0Y 498 TPSSAASDHANYAVKPEKPEKPTLASPTLAS 527
0Y 551 TASEMOQEBMKGSVEEKPBBPKRKVDITVS 580

```



```

QY 359 ED-----OGIKRIEKAANPSGKKLKIPOVVEARCA-PK 393
DB 3314 EESRRESVAERKSPLAYKEASPPASVAESI KDBAEKSESRRESVAERKSPLASKEASPT 3373
QY 394 CIGCGSSVAOPDSVYCSNDCIL-----KHAATWPLSGKEOKTKPEKVKYTKBEK 446
DB 3374 SVASVNDDEAKSKESRSDVAERKSPLASKEASPPASVAESVQDEAKSKESRRESVA 3433
QY 447 FSLPKCSVQVQIKISSVHKRLA--SEKRENPKVKMLASRSETSGKACASSTTSMASD 504
DB 3434 EKSPLASKEAS-RASVASVAYDDEAKSKESRRESVAERKSPLASKEASRPASVASVD 3492
QY 505 HNNYNAVPEKPEKPTALSPILLSKCTYHP 533
DB 3493 EAERKSESRRESVAERKSP-LPSKEASRP 3520

```

RESULT 12

T29757

protein UNC-89 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C/Accession: T29757

R/Du, Z.; Le, T.T.; Wilson, R.

Submitted to the EMBL Data Library, May 1997

A/Description: The sequence of C. elegans coemid C09D1.

A/Reference number: Z20679

A/Accession: T29757

A/Molecule type: DNA

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Residues: 1-6642 <DUZ>

A/Cross-references: EMBL:AF003131; PIDN:AB54132.1; GSPDB:GN00019; CESP:unc-89

A/Experimental source: Btiran Bristol N2; Clone C09D1

C/Genetics:

A/Map position: 1

A/Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6

3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 5.1%; Score 165; DB 2; Length 6642;

Best local similarity 19.9%; Pred. No. 0.65;

Matches 123; Conservative 87; Mismatches 214; Indels 194; Gaps 24;

```

QY 8 SNEEAPAIKPTSEKFRKTGFRRTTIKRGAGDTE--ADPSRQPOQHNLISLRSGR 64
DB 1358 ABEELPKEVINDSRKKKS---DPPDKERKSPKTEKRPASPTKGTGEVKSPEKSPA 1413
QY 65 QPKTER---VEEPLTVRRRKKXNVPLSEDS--SEPTSTVTQVE-----TASGGS 112
DB 1414 SPTCKEKSPABEYKSPKTKERKSPSSPTKCKESPSPTKGTGDEVKESPKKSPKTKES 1473
QY 113 -----VSSSSIRSGPVSLSLGEKHPASSEKAKGBEEDT-- 148
DB 1474 PEKEDVKS PVKCKEKSPDATNIVVSSSETTTEKTETTTTEHSSBSRFSVKKEKTPR 1533
QY 149 -----SDSDSDGLTAKELQNRKRRKGEQEPVRS-----LRG 180
DB 1534 KVDERPKSPPTKDKSPKSKITEIKSPVKCKESPEKVEKRPASPTKCKESPEKSPKTK 1593
QY 181 SONLRKRRREDSAGETSGVOIGSAEODRPLCKOPEASQGPVS-----QSETDI 231
DB 1594 SENEVKSPKTKERKSPKESVVE---ELKSPKESPEKADDPKSPKTKERKSPKESATDV 1649
QY 232 ENQLEKATQNTENPREACK-----PKPECEVYDNLALYICICROPHNNRPMICCDRC 285
DB 1650 KSPKTKERKSPKVEKESPTKSSPTKKTDDDEVKSPKCKG---KSP----- 1694
QY 286 EEWFGDGVGISEARGLLENGEDYICPNCTIIQVQDETNGSATNEDSGCRSVAAGDT 345
DB 1695 -----QTVBEKASPTKCKESPEKSPK----- 1715
QY 346 DCTSIGTVEOKSGBDQIKGRIEKAANPSGKKL--KIFQVVEAAGAPKICGPGCSVA 403

```

```

DB 1716 -----VEEVKSPKESPEKABKRPKSPKTKERKSPKSAABEVKSP-TKESKSPKSAE 1768
QY 404 QPDSVYCSNDCILKHAATWPLSGKEOKTKPEKVKYTKBEKPLPKCSVQVQIKISSV 463
DB 1769 KPKSPKTKESRSPVQADDEVK---SPYKCKESPEKVEKSPAS-----PTKKEKT 1814
QY 464 HKRLASREKRNPKVKMLASRSETSGKACASSTTSMASD--EKPTEL 521
DB 1815 PEKSAEELKSPKTK-----EKSPSPKTKGTGDEVKES-----PEKREKPK-- 1857
QY 522 SPTLISKCTYHPKAGFP 539
DB 1858 SPT-----PKSPPG 1867

```

RESULT 13

T21430

hypothetical protein F26H11.3b - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000

C/Accession: T21430

R/Barlow, K.

Submitted to the EMBL Data Library, November 1996

A/Reference number: Z19421

A/Accession: T21430

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-510 <WLI>

A/Cross-references: EMBL:Z81515; PIDN:CA804195.1; GSPDB:GN00020; CESP:F26H11.3b

A/Experimental source: clone F26H11

C/Genetics:

A/Map position: 2

A/Introns: 116/1; 147/3; 288/2; 348/2; 392/1

C/Superfamily: bromodomain homology

F:371-426/Domain: bromodomain homology <BRO>

Query Match 5.1%; Score 164.5; DB 2; Length 510;

Best local similarity 22.6%; Pred. No. 0.041;

Matches 70; Conservative 45; Mismatches 100; Indels 95; Gaps 14;

```

QY 94 SSEPTSTVTDTVETASBGVSSEISIRSGP-----VSDLSGKEHPASSEKAKGE 143
DB 76 SSDKTTPIKRI-IVQGPV---PIGKPRRRRCADREISEL---AAKPAAYVK 125
QY 144 EEDTSDSDGLTKEIQNRKRRKGEQEPVRSIRSGQNRKRRREDSAGTSGVOIG 203
DB 126 EVINPADITLGGDTVDYV-----KEQKPYE-SIATVSRRRRT-----SAILS 167
QY 204 SABODRPLCKB---PEASQGPVSQSETDILENQLGKATQGN-----TEENPREAG 252
DB 168 KSBODRDPKESQSTAPSKKERKTSBPASHVAFTPGSATPHDNLSTEHCTQKIPDS 227
QY 253 KPEKCE-----VYDNLALYICICROPHNN- 276
DB 228 KLYIQCELCANWYHGDGVAGVAGQTLIGLEHWSCEBCLBEGSRVADQPLVYCQKQPYDT 287
QY 277 RPMICDRCSEWPHGDCVGISEARGLLENGEDYICPNCTIIQVQDETNGSATNEDS 335
DB 288 KFYVGCSCQGMFHPBCVGTTRABA---EQADYNCPACT-----RAABGYESASDVS 338
QY 336 GCRSVAGDGT 345
DB 339 GSSRSVQOLT 348

```

RESULT 14

I51618

nucleolar phosphoprotein - African clawed frog

C/Species: Xenopus laevis (African clawed frog)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999

C/Accession: I51618; S57757

R/Cairns, C.; McStay, B.

J. Cell Sci. 108, 3339-3347, 1995
 A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
 A:Reference number: 151618; MUID:96019267; PMID:7593294
 A:Accession: 151618
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-390 <CAI>
 A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
 C:Genetics:
 A:Gene: XNopp180
 C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
 C:Keywords: phosphoprotein

Query Match 5.1%; Score 163.5; DB 2; Length 990;
 Best Local Similarity 18.8%; Pred. No. 0.098;
 Matches 128; Conservative 78; Mismatches 248; Indels 227; Gaps 24;

3 DKGHLSNEB-----APYAIPTSEKFRKTNQFRRTTIKREAGDTEADPS 48
 297 DSSDSDSEBQPAKAKIYPAKAAASAPKPLAK-----KATSTDSSESDS 342

49 EQQPQAHNLISLRSGRPRTERTVERFLTVRRGKGVPSLEDSSEPTSTVTVDERA 108
 343 SEDEKSSVVLGVKA-APKKAAPAPAKSTPVAAAKKSAAPAKKASSSDSDSSNEETTT 401

109 SSGSVSSSEIRSGPV--DSLGKEHPASSRAKAG-----BEEEDTSDSDSD-G 155
 402 KPAKTTTAKSAATPTSKTPTMGKATPTSKTPAKPTPTSTAKDSSSDSDSDSD 461

156 LTKELQNLRRRBOEPVERSLRGSQNLRRK-----REEDSAETGVSQIG 203
 462 TTKPAKTTAKSAATPTSKTPTMSKATPTSKTPTAKPTPTSAKDSSDSDSDSDS 521

204 SAEOD-----RPLKQEP-----BASQGPVQ 225
 522 SDEKTPAKRAKATTPAKPAKTTTAPKPAKTTTAPKPAKSTPGKVPPTKSSSSSDSD 581

226 SETDDIENQLBG---KATQNTENPRBAGKPECEVYDPNALYICIQPHNRRFMIC 282
 582 SSSSEBKSSAPAKTTTGKATSKGVVASKVPAAKASSSD-----S 625

283 DRCEEMFHGDVYGISBARGRLLENGEDYICPNTIL---QVODETNGSATNEQDSCRS 339
 626 DSSBEETTKTKPLTK-----LSPAVKTLPPKLAESSSDSDSDSEKTKRP 672

340 VGADGTDCISIGTV-----EQSGEDGIGRIEKANPSGKKLKIPOPV 385
 673 AKPPAKSATPVNTKAPAKOKRAKASCDSDDSSSEBEG-----SKQPTGK-----SPA 720

366 VEPAPGAPKICIGCGSSVVAQPDVYCSNDCILKHAATMRFLSSGKEQKTPKKEVTKPE 445
 721 AKATAPPK-KNPPVAVKPKDPTSSSSSD-----SSGDBKOKPKQAAAAKDV 765

446 KFS-----LPRCSVQVQIGIKISSVHKRLASEKREN-----PV-KKVMLASSEETSG-- 489
 766 KQGAKAAPKTPKKAASSSESDSDSDVSKAKKTNTAVSKSPVPTTPKAVPAKKESSSES 825

490 -----KEAACSSSTPSMASDH-----NY 507
 826 SDEDEKQGGKNTSTTKIANSTPKAAACSBESSSSBEGKANGTSGRRKRSTGNABC 885

508 NAVKPEKPKPTALSPFTLLSK 528
 886 EAVTPEN-KGLKAKSPNTFPK 905

RESULT 15
 T18837
 hypothetical protein C01G6.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18837
 R:Berts, M.

submitted to the EMBL Data Library, August 1994
 A:Reference number: Z19029
 A:Accession: T18837
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-952 <ML>
 A:Cross-references: EMBL:Z35595; PIDN:CAA84636.1; GSPDB:GN00020; CESP:C01G6.5
 C:Genetics:
 A:Gene: CESP:C01G6.5
 A:Map position: 2
 A:Introns: 34/2; 61/2; 101/3; 195/1; 218/3; 845/3; 886/3

Query Match 5.0%; Score 161.5; DB 2; Length 952;
 Best Local Similarity 23.9%; Pred. No. 0.12;
 Matches 60; Conservative 41; Mismatches 111; Indels 39; Gaps 9;

90 SLEDSSEPTSTTVTVETASBGVSSESSRI-----RSGPVSDSLGKEHPASSRK 138
 678 SVAREEETSSKENTDQKRAVDDSDSDSARLVKQIKALSATPSSTPASSRKQKOTSSRK 737

139 AKGGEHEEDTSDSDGLTLKELQNLRRRBOEPVERSLRGSQNLRRKRRREDSATG 198
 738 KQQLDSSADSDSDSDGREGDKSTKRDKARSTRIGKTPAKVG-KKAKYEDBEDTD 796

199 SVQIGSAQDRPLCK-----QEPASQGPVSGSRT--DIEHQLEKATQANTENPRE 250
 797 DVQ-----EERKVTPRGRRKRMKRTASQAIKERKTKDQVBEBAPKKRGRKATPTE 852

251 AGKP-----KPECEYDPRALYICIR--QPHNRRFMICDRCEEMFHGDVYGISBARG 302
 853 EEEPPPTBPSKRCGVAKG---HCISAKYERKRLQWVSCSICNQMFHWCVRLD---N 905

303 LLENGEDYIC 313
 906 VCTREDETEFLC 916

Search completed: April 20, 2004, 21:20:24
 Job time : 24 secs

CC -1- FUNCTION: Putative transcription factor, weakly pro-apoptotic when overexpressed.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus after pro-apoptotic stimuli.
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Expressed at intermediate levels.
 CC -1- INDUCTION: Upregulated during apoptosis.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.emb-emb.ch/announce/or_send_an_email_to_license@emb-emb.ch).
 CC -----
 CC EMBL; AJ238333; CAB48401.1; -
 CC EMBL; AK042474; BAC31270.1; -
 CC EMBL; AK044919; BAC32141.1; -
 CC MGI; MGI:1344352; D1d01.
 CC GO; GO:0005634; C:nucleus; IDA.
 CC GO; GO:0006915; P:apoptosis; IDA.
 CC InterPro; IPR001965; Znf_PHD.
 CC Pfam; PF00628; PHD; 1.
 CC SMART; SM00249; PHD; 1.
 CC PROSITE; PS01359; ZF_PHD_1; 1.
 CC PROSITE; PS0016; ZF_PHD_2; 1.
 CC Apoptosis; Nuclear protein; Zinc-finger.
 CC DOMAIN 162 170 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC ZN_FING 182 190 PHD-TYPE.
 CC FT 285 319 V -> A (IN REF. 1).
 CC FT 45 45 D -> N (IN REF. 1).
 CC FT 331 331 I -> V (IN REF. 1).
 CC FT 353 353
 CC SEQUENCE 614 AA; 67409 MW; 15ACCS5A8ABF3539 CRC64;

Query Match 99.7%; Score 3218; DB 1; Length 614;
 Best Local Similarity 99.5%; Pred. No. 2,4e-165;
 Matches 611; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDDKGLSNEAPKAIKPTSKFRRTIAKRGAGTADPSEQPOQHNLISLR 60
 DB 1 MDDKGLSNEAPKAIKPTSKFRRTIAKRGAGTADPSEQPOQHNLISLR 60
 QY 61 RSGROPKTERVEEFLTYRRRKKKNVPVLSDESSPTSTVTVDETASGVSSSSBIR 120
 DB 61 RSGROPKTERVEEFLTYRRRKKKNVPVLSDESSPTSTVTVDETASGVSSSSBIR 120
 QY 121 SGFVDSLSKKEPSSERKAGGEBEDTSDSDGLTLKELQNRLLRRKEQEPVERSLRG 180
 DB 121 SGFVDSLSKKEPSSERKAGGEBEDTSDSDGLTLKELQNRLLRRKEQEPVERSLRG 180
 QY 181 SONRLRRKRREDSAYTSVQISAGQDRPLCKQEPASQSPVSOSETDINOLEGKAT 240
 DB 181 SONRLRRKRREDSAYTSVQISAGQDRPLCKQEPASQSPVSOSETDINOLEGKAT 240
 QY 241 QGNTENPREACKPRCECVDPNMLYCI CRQPHNNRPMICCDRCCEWFGDCVGISBAR 300
 DB 241 QGNTENPREACKPRCECVDPNMLYCI CRQPHNNRPMICCDRCCEWFGDCVGISBAR 300
 QY 301 GRLLRNEDYICNPCTTIQVNDENGTNTNODSCGRCVAGDGTCTSIGVVEQSGSD 360
 DB 301 GRLLRNEDYICNPCTTIQVNDENGTNTNODSCGRCVAGDGTCTSIGVVEQSGSD 360
 QY 361 GILLRNEDYICNPCTTIQVNDENGTNTNODSCGRCVAGDGTCTSIGVVEQSGSD 360
 DB 361 GILLRNEDYICNPCTTIQVNDENGTNTNODSCGRCVAGDGTCTSIGVVEQSGSD 360
 QY 421 ATWRFLSSGKEQKTKPEKVKTKPKESIPKCSVOVGICSSVHKKLASKEKNPVYKTM 480
 DB 421 ATWRFLSSGKEQKTKPEKVKTKPKESIPKCSVOVGICSSVHKKLASKEKNPVYKTM 480
 QY 481 LASRBTSGKEAKCSSSTPSMASDHNHYNAVKEPKERPALAPTLLSKCTYHPKAGFP 540

DB 481 LASRBTSGKEAKCSSSTPSMASDHNHYNAVKEPKERPALAPTLLSKCTYHPKAGFP 540
 QY 541 SHRLGCGGLSTRVILVLIIVASSSLPARSRVYDASGPVFLPSLNSLGMFLKSCVG 600
 DB 541 SHRLGCGGLSTRVILVLIIVASSSLPARSRVYDASGPVFLPSLNSLGMFLKSCVG 600
 QY 601 LMLRALSYSPRP 614
 DB 601 LMLRALSYSPRP 614
 RESULT 2
 ID DAT1_HUMAN STANDARD; PRT; 1225 AA.
 AC Q9BHC0; O15043; Q9BHC0; Q9BHC0; Q9H4G6; Q9H4G7; Q9NTU8; Q9NTU8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Death associated transcription factor 1.
 GN DAT1 OR KIA0333.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosofri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takehashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saio K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuda Y.,
 RA Niimura K., Iwayanagi T.;
 RT NEDO human cDNA sequencing project.
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
 RA Buck D., Burrill M.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.B.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle B., Hunt A.R., Hunt S.B., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA LeHayesalho M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.P., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.B., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Whitehead L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20."
 RL Nature 414:865-871 (2001).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2 AND 3).
 RC TISSUE=Colon, Kidney, and Muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Bietow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datschenko L., Marinska K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Pabey J., Helton B., Keltman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Gilmwood J., Schmitz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalins D.B.,
RA Schnerch A., Schein J.R., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences." Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RP SEQUENCE OF 235-1225 FROM N.A. (ISOFORM 1).
RT TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nsague T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -1- FUNCTION: Putative transcription factor, weakly pro-apoptotic when
overexpressed (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus
after pro-apoptotic stimuli (by similarity).
CC -1- ALTERNATIVE PRODUCTS:
Brent-Alternative splicing; Named isoforms-3;
Name=1;
IsoId=Q9BTC0-1; Sequence=Displayed;
Name=2;
IsoId=Q9BTC0-2; Sequence=VSP_007208, VSP_007209, VSP_007210;
Note= No experimental confirmation available;
Name=3;
IsoId=Q9BTC0-3; Sequence=VSP_007208, VSP_007211, VSP_007212;
Note= No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -1- CAUTION: Ref. 2 sequence differs from that shown due to erroneous
gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AK002127; BAA92094.1; -;
CC EMBL; AL035669; CAC12755.1; ALT_SEQ.
CC EMBL; AL035669; CAC12756.1; -;
CC EMBL; AL035669; CAC28883.1; -;
CC EMBL; AL117379; CAB90275.1; -;
CC EMBL; BC000770; AAH00770.1; -;
CC EMBL; BC004237; AAH04237.1; -;
CC EMBL; BC014489; AAH14489.1; -;
CC EMBL; AB002331; BAA20791.1; -;
CC EMBL; HGNC:2680; DAF1.
CC MIM; 604140; -;
CC InterPro: IPR003618; TFS2_centre.
CC InterPro: IPR001965; Znf_PHD.
CC Pfam; PF00628; PHD; 1.
CC SMART; SM00249; PHD; 1.
CC SMART; SM00510; TFS2M; 1.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS50016; ZF_PHD_2; 1.

KW Apoptosis; Nuclear protein; Zinc-finger; Alternative splicing.
FT DOMAIN 165 173 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 168 193 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN FING 268 322 PHD-TYPE.
FT VASPLIC 388 423 Missing (in isoform 2 and isoform 3).
FT VASPLIC 566 580 /FTId=VSP_007208.
FT VASPLIC 566 580 STKDRSRSEKKAAT -> CSKIKYLSHPSLIA (in
isoform 2).
FT VASPLIC 581 1225 /FTId=VSP_007209.
FT VASPLIC 566 598 Missing (in isoform 2).
FT VASPLIC 566 598 /FTId=VSP_007210.
FT VASPLIC 566 598 STKDRSRSEKKAATVPQSGTWGKOP -> CNYH
LGVELDPSRSFWIAPACPGIQAALC (in isoform
3).
FT VASPLIC 599 1225 /FTId=VSP_007211.
FT VASPLIC 599 1225 Missing (in isoform 3).
FT VASPLIC 599 1225 /FTId=VSP_007212.
FT VASPLIC 599 1225 G -> E (IN REF. 1).
FT VASPLIC 599 1225 T -> M (IN REF. 2).
FT VASPLIC 599 1225 T -> A (IN REF. 2).
SQ SEQUENCE 1225 AA; 133023 MW; 595071B0D3C74A2 CRC64;

Query Match 63.6%; Score 2052.5; DB 1; Length 1225;
Best Local Similarity 72.2%; Pred. No. 1.1e-102;
Matches 412; Conservative 32; Mismatches 82; Indels 45; Gaps 7;

QY 1 MDKCHLSTNRBAPKAIKPTSKKFRKTWGFRTTIARREGAGTEADPSQQP--QQHNL 58
DB 1 MDKDDPSNRBAPKAIKPTSKKFRKTWGFRTTIARREGAGTEADPSQQPQQGLGHS 60
QY 59 LRRSGROPRTTTRVBEPLTTTTRRRKQAVPVLSSESTSTVTVDFVTRASGVSSSS 118
DB 61 LRRSGROPRTTTRVBEPLTTTTRRRKQAVPVLSSESTSTVTVDFVTRASGVSSSS 120
QY 119 TRSGPVDSLSG--KEBPASERKAGGEEEDTSDSDGLTLKELQNRLLRKEQBEVRS 177
DB 121 TRSGPVDSLSG--KEBPASERKAGGEEEDTSDSDGLTLKELQNRLLRKEQBEVRS 180
QY 178 LRRSGONRLRKKRRRSDSMT--GSVOIGSNAEDRPLCKQEPBASQGVSGSEFTDDIENOLE 236
DB 181 LRGISRLRKKRRRREGPATTVGSBASDVTVGVLP--SKQEPNDQGVVSGAGQDSDRSKLE 239
QY 237 GATGATGNTENPREAKPEKPECEVDPNALYCI CROPENNRPMI CDRCEBEPHGCVCVI 296
DB 240 GAAQDIDKEEKGDLARPECECEGYDPNALYCI CROPENNRPMI CDRCEBEPHGCVCVI 299
QY 297 SARARGLERNGEDYICPCTILQVODETNGSATNBODSGCSVAGDGTDCSTIGTVBOK 356
DB 300 SARARGLERNGEDYICPCTILQVODETNGSATNBODSGCSVAGDGTDCSTIGTVBOK 359
QY 357 SGEEDGIRKIRKAAKNSGKKLKIPOP----- 384
DB 360 SGEEDGIRKIRKAAKNSGKKLKIPOP----- 384
QY 385 ----VYBAPGAPKCIQPGGCSVAQDPVYCSNDCLIKRAAATRTSSGEGQTKPKYK 440
DB 420 ISCKVITBARGAKKICGPGCCHVAQPSVYCSNDCLIKRAAATRTSSGEGQTKPKYK 479
QY 441 KTKPEPESLPKCSVGVGKISSVHKLASEKRNPKK--VMLASREFTGKKAACSSSTP 499
DB 480 KMKPEPESLPKCSVGVGKISSVHKLASEKRNPKK--VMLASREFTGKKAACSSSTP 539
QY 500 SWASDHNNAVAKPEKPEPTALSPITLSKCT 530
DB 540 SWASDHNNAVAKPEKPEPTALSPITLSKCT 567

RESULT 3
PRT3_HUMAN STANDARD; PRT; 2039 AA.
AC Q92576; Q9N016; Q9N145;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)

10-OCT-2003 (Rel. 42, last annotation update)
 PHD finger protein 3.
 PHF3 OR KIA0244.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxId=9606;
 [1]
 SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 MEDLINE=21845875; PubMed=11856869;
 RA Fischer U., Struss A.-K., Hemmer D., Michel A., Henn W.,
 RA Steudel W.-I., Meese B.;
 "HPF3 expression is frequently reduced in glioma."
 RL Cytochrome. Cell Gene. 94:131-136(2001).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=bone marrow;
 RC MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
 RA Ohara O., Tanaka A., Kocant H., Miyajima N., Komura N.;
 "Prediction of the coding sequences of unidentified human genes. VI.
 The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain."
 RT DNA Res. 3:321-329(1996).
 [3]
 SEQUENCE OF 83-2039 FROM N.A.
 RA Matthews L.;
 RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Expression is significantly
 CC reduced or lost in glioblastoma, glioblastoma cell lines,
 CC anaplastic astrocytomas, and astrocytomas.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL: AF091622; AAF21292.1; -
 DR EMBL: D87685; BA013438.2; ALT_INTR.
 DR EMBL: AL050339; CAB9358.1; -
 DR Genew; HGNC:8921; PHF3.
 DR MIM: 607789; -
 DR InterPro; IPR003618; TFS2 centre.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF00628; PHD; 1.
 DR SMART; SM00249; PHD; 1.
 DR SMART; SM00510; TFS2M; 1.
 DR PROSITE; PS01359; ZP_PHD_1; 1.
 DR PROSITE; PS0016; ZP_PHD_2; 1.
 KM Zinc-finger.
 FT ZN FING 717 772 PHD-TYPE.
 FT DOMAIN 929 1029 TFS2M.
 FT DOMAIN 1797 1865 PRO-RICH.
 SQ SEQUENCE 2039 AA; 229480 MW; PDAFF005760059B CRC64;
 Query March 7.4%; Score 239.5; DB 1; Length 2039;
 Best local similarity 20.5%; Pred. No. 1.9e-05;
 Matches 101; Conservative 75; Mismatches 163; Indels 153; Gaps 17;
 QY 90 SLEDSSEPTSTVDVETASRGVSSEIRSGVSDSLGK-----EHPASSKA 139
 DB 466 NLQDDRNQSSSVLSKSKVSKTKKPKVHSKQMTTDAKIVAAKYEVIHSKTKVNV 525
 QY 140 KGEHEERTSDDSGLTAKLQV-----RLRKR-EGEPV-----ERSLRSGQNLIR 186
 DB 526 KSVKRLTVDVPSQ-----QNFHPRVKRKQIDEPKIQSGNSGVKSVKQNAHSVL 576
 QY 187 KKRREDSAEFGSVQI-----GSAQDPRPLCKQEP--ASQGVQSGE----- 227

577 KRTLDQDTL-----VQIFKPLTHLSLSDKSHAHGCLKPRHPAQTGHVSHS0KQCKRPQ 632
 QY 228 -----TDIENQLEGKATQNTENRPEACKPRPCE----- 259
 DB 633 QAPAKNTSHVEBLEHFGVHFKEEDKTKLKPKPKKNQOPRORSSKSESLDEPLPTD 692
 QY 260 -----VYDPNALYICIGPFPNNFMICCDRCBEYFHGDGVGISBA 299
 DB 693 NIATTRRGSDHSSSPBSKRTMTPEKQGFCKPHGNTFMVCGKCDMDPHDGVGLSL 752
 QY 300 RGRLLRNEEDYICPNC-----TILQVDETNGSATNEODSGRSVGADTCTSIGT 352
 DB 753 QAQCGEBDKYCVKCCABDKTEILDPITLQVATVEFISGDKTM-----ECKKGL 807
 QY 353 VEQKSGE-----DQIKRIEKAANPSGKKLQPVVVEAFGARPCIGPGCCSVAPQPS 407
 DB 808 SKHTTNDRTKYIDTVKVKVILKRESGSR-----NS 840
 QY 408 VYCSNDCLKHAATRTLSGKQKTPKPKVTKPRPSLPRCSVOYGIKISVHRL 467
 DB 841 SDCRNEIKKQOLAPLRMG-----QVLPKSSSEKSEK--IPKSTVYTCGERASFG 894
 QY 468 ASEKENPVKCV 479
 DB 895 THEKQEMKKCV 906
 RESULT 4
 ID_NEST RAT STANDARD; PRT; 1805 AA.
 AC P21263;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nestin.
 GN NES.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OK NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90150286; PubMed=1689217;
 RA Lendahl U., Zimmermann L.B., McKay R.D.G.;
 RT "CNS stem cells express a new class of intermediate filament
 RT protein."
 RL Cell 60:585-595(1990).
 CC -1- TISSUE SPECIFICITY: CNS STEM CELLS.
 CC -1- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
 CC IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
 CC -1- SIMILARITY: Belongs to the intermediate filament family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL: M34384; AAA41685.1; -
 DR PIR: A34736; A34736.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; Filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KM Intermediate filament; Coiled coil; Neurone.
 FT DOMAIN 1 7 HEAD.
 FT DOMAIN 8 314 ROD.
 FT DOMAIN 315 1805 TAIL.
 FT DOMAIN 8 43 COIL 1A.
 FT DOMAIN 44 55 LINKER 1.
 FT DOMAIN 56 151 COIL 1B.
 FT DOMAIN 152 174 LINKER 12.


```

QY 220 QGPVSGSTDDIENOLEGKATQNTENPREAKRPECEVYDPAALYICRQPHNNFM 279
DR 1197 QADITSSSDIFDDQNSIGESSDE---QKIKVTENLVLSHTGPC----- 1242
QY 280 ICCRCCEWFGDCCVGISBAGRL-----BRNEDVICRCTLQVQ---DFTN 326
DB 1243 -----QSSGD-----BALSKSVPTVDDDDDDNDENNAKQMLBETKANISBED 1289
QY 327 GSATNEODSGCRSVAGDGTCTSIGTVEQKSGEDGIGRIEKAAMPKGGKLIKIPQPV 386
DB 1290 GSSDDEPEBEGKRRKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1333
QY 387 EAPAPKICIGGCCSSVAPDSVYCSNDCTILKHAATMFLSSGKQKTKPKK---KVKTKP 444
DB 1334 -----PR-----YRHLRLRH-KLTVSDGSGEGEKTKPKKPKKQKQKQKRN 1370
QY 445 EKFLSPKCSVQVGIKISVHKRLASKEKRNPKYKMLASRS-----TSQKRAACESSTP 499
DB 1371 RR-----KVSSBDBE-DSDPQSGVSESVSESEDEQRPRTSAKKALEENOR 1417
QY 500 SM-----ASDHNYNAVPEKPEK 517
DB 1418 SYQKKKRRRIKQVDESSSENKSNSEBEEK 1449

RESULT 6
ID ATRX CAEEL STANDARD; PRT; 1359 AA.
AC Q9UT80; Q02061;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetidae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=99365296; PubMed=10433961;
RA Villard L., Fontes M., Ewbank J.J.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT the human XNP/ATR-X gene."
RL Gene 236:13-19(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton R., Wohlmann P.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Could be a global transcriptional regulator. Modifies
CC gene expression by affecting chromatin (Potential).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.emb-net.ch/announce/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: AF134186; AAD55361.1;
CC EMBL: AF000196; AAC24256.1;
CC PIR: T34036; T34036.
CC WormPep; B0041.7; CE17314.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR000330; SNF2_N.
CC Pfam; PF000271; helicase_C_1.
CC Pfam; PF00176; SNF2_N_1.
CC SMART; SM00487; DEXDC; 1.

```

```

DR SMART; SM00487; HELIC; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; FALSE_NEG.
KM DNA repair; Hydrolyase; Helicase; Nuclear protein; ATP-binding;
KM DNA-binding.
FT NP BIND 496 503
FT SITE 636 639
FT DOMAIN 67 70
FT DOMAIN 266 272
FT DOMAIN 276 281
FT DOMAIN 372 375
FT DOMAIN 603 608
FT DOMAIN 859 862
FT COMPLET 479 479
SQ SEQUENCE 1359 AA; 156191 MW; BB4342547D4F64 CRC64;

Query Match 5.7%; Score 185; DB 1; Length 1359;
Best Local Similarity 21.4%; Pred. No. 0.0098;
Matches 146; Conservative 88; Mismatches 217; Indels 232; Gaps 33;

QY 8 SNEAPKAIKPTSKREPKRTWGRRTTIKRBAGDTPADPSE--QDPQHNLRLRSGRQ 65
DB 210 SESEDEKVKKKKKKKKKK-----VVKSESSEDEAPKKKTKRKRSKTSSEBSSES 261
QY 66 PRTTREVREPLTVVRBGRKQVPSLSDSEPTSTVTVETASE-----GSV-----ESSS 117
DB 262 EKSDDEEEKESSPPKPKKK--PLAVKGLSDSESEBBDVRLPQKKRGAVTLISDSED 319
QY 118 EIRSGVSDSLQKHPASSERAKGEEBEDTSDSDGLITLKLONLRARRKQ--EPVER 176
DB 320 EKDQKSEBSADVEKVKAKKQSSSESGDSSEGSITV---NRKSKKEKKEKKKK 375
QY 177 SLRSGQNLKRRKREDEDAFTGVSIGSABORPLCKQEPKASQGPVSGSTDDIENOLE 236
DB 376 GLINDSSKLQKETTDAERAE-----KERRKRLKKQKERN--GIYLBESDLEMLT 425
QY 237 GKATQNTRE---NP-----REAGKPRPECEVYDPAALYICR--QPHNRFMTC--- 282
DB 426 GTSSGRKLKSVLDDSDSTVDESKKP---VEVH--NLVRLTKPHQAHGIFMWDCAE 480
QY 283 --DRCEWFGH--DCVGISBAGRL-----BRNED-----YICPCTIL 319
DB 481 SLDRLOTGSGGILAHCMGL-----GKTLQVITLHTVLMHKEKIGCKKRVLVVPPQVIT 536
QY 320 -----QVQPEFGSATNBOD-----SGCRSVAGDGTCTSIG 351
DB 537 NWFKEFKQVLDNDEELTITVNEIDSYKTIEDRRALKAMWSKTPSMITGYDLFRL 596
QY 352 TYEQ--KSGEDGIGRIEKAAMPKGGKLIKIPQVVEAPGAPKICIGPCSSVAPDSVY 409
DB 597 TVEDDPKKKKKPKRRRLKAKED-----FRKYLQNP-----PDVV 634
QY 410 CSNDCTILKHAATMFLSSGKQKTKPKKPKKQKTSPLKCSVQVGIKISVHKRLAS 469
DB 635 CD-----EAHKLKNDLSLISKQNYI----- 655
QY 470 EKRBNPKYKMLASSTSGKRAACSSSTPMSADHNTVAVAPKPEKPTALPTLLSKC 529
DB 656 -----LTKRRLITLTP-----LQNNIME--- 674
QY 530 TYHPKAGFGPSHILGCGISLRTVLGLVLIASSSLPARRYQDASGPV--FLPSLW 588
DB 675 -YHCANVFKP-----GLLG--TKTEPARNFVMI-----NGRTVDAPLFVSEFKRR 721
QY 589 SLSGWFLNSCVG-----LMLENI 606
DB 722 HVLVDHLKKCVDRKQYRVLTETAI 744

RESULT 7
ID ATRX HUMAN STANDARD; PRT; 2492 AA.
AC P46100; P51068; Q15886; Q9H021; Q9NTS3;
DT 01-NOV-1995 (Rel. 32, Created)

```

DT 28-FEB-2003 (Ref. 41, Last sequence update)
DT 15-MAR-2004 (Ref. 43, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOPRIMS 1; 2; 3; 4 AND 5), VARIANTS SER-596;
RP GLU-740 AND SER-1860, AND VARIANTS ATR-X.
RA MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J.; Higgs D.R.; Bachoo S.; Blake D.J.; Quarrell O.W.J.;
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RT Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOPRIMS 2 AND 4).
RA MEDLINE=97386582; PubMed=9244431;
RA Villard L.; Losel A.-M.; Cardoso C.; Proud V.; Chiaroni F.;
RA Colleaux L.; Schwartz C.; Fontes M.;
RA "Detection of the genomic structure of the XNP/ATRX gene encoding
RA a potential zinc finger helicase.";
RA Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE FROM N.A. AND VARIANTS SER-596 AND GLU-740.
RA MEDLINE=22763540; PubMed=1277553;
RA Kitzano T.; Schwarz C.; Nickel B.; Paabo S.;
RA "Gene diversity patterns at 10 X-chromosomal loci in humans and
RA chimpanzees.";
RA Mol. Biol. Evol. 20:1281-1289(2003).
RN [4]
RP SEQUENCE OF 860-2492 FROM N.A.
RA MEDLINE=95179111; PubMed=7874112;
RA Stayton C.L.; Dabovic B.; Gulsano M.; Gecz J.; Broccoli V.;
RA Giovannazzi S.; Bossoiaco M.; Monaco L.; Raetan S.; Boncinelli B.;
RA Bianchi M.B.; Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RT Hum. Mol. Genet. 3:1957-1964(1994).
RN [5]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RA MEDLINE=94214473; PubMed=8162050;
RA Gecz J.; Pollard H.; Consalez G.; Villard L.; Stayton C.L.;
RA Milasseau P.; Kiretschatsky M.; Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RT Hum. Mol. Genet. 3:39-44(1994).
RN [6]
RP SEQUENCE OF 2401-2492 FROM N.A. AND VARIANTS ATR-X.
RA MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J.; Picketts D.J.; Villard L.; Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RT Cell 80:837-845(1995).
RN [7]
RP SEQUENCE OF 1375-2492 FROM N.A.
RA Pearce A.; Chapman J.;
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [8]
RP BZHZ BINDING.
RA MEDLINE=98167853; PubMed=9499421;
RA Cardoso C.; Timelst S.; Villard L.; Kiretschatsky M.; Fontes M.;
RA Colleaux L.;
RT "Specific interaction between the XNP/ATR-X gene product and the SER
RT domain of the human BZHZ protein.";
RT Hum. Mol. Genet. 7:679-684(1998).
RN [9]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RA MEDLINE=20040663; PubMed=10570185;

RA McDowell T.L.; Gibbons R.J.; Sutherland H.; O'Rourke D.M.;
RA Bickmore W.A.; Pombo A.; Hurley H.; Gatter K.; Picketts D.J.;
RA Buckle V.J.; Chapman L.; Rhodes D.; Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATRX) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN [10]
RP DISEASE.
RA MEDLINE=20213147; PubMed=10751095;
RA Villard L.; Fontes M.; Ades L.C.; Gecz J.;
RT "Identification of a mutation in the XNP/ATR-X gene in a family
RT reported as Smith-Petersen-Myers syndrome.";
RT Am. J. Med. Genet. 91:83-85(2000).
RN [11]
RP DISEASE.
RA MEDLINE=22804448; PubMed=12858175;
RA Gibbons R.J.; Pellagatti A.; Garrick D.; Wood W.G.; Malik N.;
RA Ayub H.; Langford C.; Boulwood J.; Waincoat J.S.; Higgs D.R.;
RT "Identification of acquired somatic mutations in the gene encoding
RT chromatin-remodeling factor ATRX in the alpha-thalassemia
RT myelodysplasia syndrome (ATMDS).";
RT Nat. Genet. 34:446-449(2003).
RN [12]
RP VARIANT ATR-X SER-1713.
RA MEDLINE=97196774; PubMed=9043863;
RA Villard L.; Lacombe D.; Fontes M.;
RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
RT without alpha-thalassemia.";
RT Eur. J. Hum. Genet. 4:316-320(1996).
RN [13]
RP VARIANT JM GLN-2131.
RA MEDLINE=96224392; PubMed=8630485;
RA Villard L.; Gecz J.; Mattei J.-F.; Fontes M.; Saugier-Veber P.;
RA Munich A.; Lyonnet S.;
RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";
RT Nat. Genet. 12:359-360(1996).
RN [14]
RP VARIANTS ATR-X.
RA MEDLINE=97467722; PubMed=9326931;
RA Gibbons R.J.; Bachoo S.; Picketts D.J.; Aftimos S.; Aisenbauer B.;
RA Bergoffen J.; Berry S.A.; Dahl N.; Fryer A.; Keppler R.; Kuroawa K.;
RA Levin M.L.; Masuno M.; Neri G.; Pierpont M.E.; Slaney S.F.;
RA Higgs D.R.;
RT "Mutations in transcriptional regulator ATRX establish the functional
RT significance of a PHD-like domain.";
RT Nat. Genet. 17:146-148(1997).
RN [15]
RP VARIANT ATR-X LEU-246.
RA MEDLINE=20123062; PubMed=10660327;
RA Fichera M.; Romano C.; Castiglia L.; Palla P.; Ruberto C.; Amata S.;
RA Greco D.; Cardoso C.; Fontes M.; Ragnua A.;
RT "New mutations in XNP/ATR-X gene: a further contribution to
RT genotype/phenotype relationship in ATR/X syndrome.";
RT Hum. Mutat. 12:214-214(1998).
RN [16]
RP VARIANT SHS LYS-1742.
RA MEDLINE=99347960; PubMed=10417298;
RA Losel A.-M.; Millan J.M.; Villard L.; Orellana C.; Cardoso C.;
RA Prieto F.; Fontes M.; Martinez F.;
RT "Mutation of the XNP/ATR-X gene in a family with severe mental
RT retardation, spastic paraplegia and skewed pattern of X inactivation:
RT demonstration that the mutation is involved in the inactivation
RT bias.";
RT Am. J. Hum. Genet. 65:558-562(1999).
RN [17]
RP VARIANT CWS THR-2050.
RA MEDLINE=99326061; PubMed=10398237;
RA Abidi F.; Schwartz C.B.; Carpenter N.J.; Villard L.; Fontes M.;
RA Curtis M.;
RT "Carpenter-Maziri syndrome results from a mutation in XNP.";
RT Am. J. Med. Genet. 85:249-251(1999).
RN [18]

RP VARIANTS ATR-X GLU-175; 178-VAL--LYS-198 DEL; SER-190; PRO-219;
 RP LBU-246 AND CYS-249;
 RX MEDLINE=99219535; PubMed=10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belongue J.,
 RA Loefer A.-M., Seaver L., Bonnefont J.-P., Romano C., Ficheria M.,
 RA Lecomte D., Hanauer A., Philip N., Schwartz C.B., Pons M.,
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 ATR-X syndrome";
 RT J. Med. Genet. 36:183-186(1999).
 RN (19)
 RP VARIANTS ATR-X SER-179; LBU-190; ILR-194; CYS-246; PHE-1552; SER-1645
 RP AND CYS-1847.
 RX MEDLINE=20451413; PubMed=10995512;
 RA Mada T., Kubota T., Fukushima Y., Saitoh S.;
 RT "Molecular genetic study of Japanese patients with X-linked alpha-
 thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 RN (20)
 RP VARIANT ATR-X MET-1621.
 RX MEDLINE=22108790; PubMed=12116232;
 RA Yntema H.G., Poppelars P.A., Derksen E., Oudakker A.R.,
 RA van Roozemaalen T., Jacobs A., Obbema H., Brunner H.G., Hamel B.C.J.,
 RA van Bokhoven H.;
 RT "Expanding phenotype of XNP mutations: mild to moderate mental
 retardation";
 RT Am. J. Med. Genet. 110:243-247(2002).
 CC -1- FUNCTION: Could be a global transcriptional regulator. Modifies
 gene expression by affecting chromatin. May be involved in brain
 development and facial morphogenesis.
 CC -1- SUBUNIT: Probably binds E2H2. Binds annexin V in a calcium and
 phosphatidylcholine/phosphatidylethanolamine-dependent manner (By
 similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
 heterochromatin during interphase and mitosis, probably by
 interacting with HPI.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing. Named isoforms=5;
 CC Name=4;
 CC IsoId=P46100-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=P46100-2; Sequence=VSP_000575;
 CC Name=2;
 CC IsoId=P46100-3; Sequence=VSP_000574;
 CC Name=3;
 CC IsoId=P46100-4; Sequence=VSP_000576;
 CC Name=5;
 CC IsoId=P46100-5; Sequence=VSP_000574, VSP_000576;
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- DISEASE: Defects in ATRX are the cause of X-linked alpha-
 thalassemia/mental retardation syndrome (ATR-X) [MIM:301040]. ATR-
 X is an X-linked disorder comprising severe psychomotor
 Query Match 5.7%; Score 184.5; DB 1; Length 2492;
 Best Local Similarity 18.7%; Pred. No. 0.02;
 Matches 107; Conservative 113; Mismatches 201; Indels 151; Gaps 24;
 QY 3 DKGLSHNEAPATPTSTSEKPKTWTGPRRTTAKRGGAGDTADSEQPOGHNLSTARS 62
 DB 972 DQSDSTSDDDKQKQKGTTEKKVPSDFKKVY-KHQEQYESSDGTETKPKREKIC- 1026
 QY 63 GROPRTREVEPLTTRRRGKKKPVSLDESSEPTSTVTVEYASBG---SVSSSR1 119
 DB 1027 -HFGKIGIKQIKGTGTDGKSKK-----IRDTSKKKKDELSDYAKSTGKGSCHSSBRX 1080
 QY 120 RSGPVSDSLGKHPASSERAKGSEEBEDTSDSDGLTLKR-----LQNLRR- 167
 DB 1081 KS--KNGAYGRKKCKKLKLGSSRRKQDCSSSDTEKYSMKSGCNSSDRLKRIELERR 1138
 QY 168 ----RREGQPVRSURGS----QNLKRRRRRSDSAFTSGVIGSEQRPPLCKQEPERS 219
 DB 1139 NLSSKRNTEKIQSGSSSSDAESSSDNNKCKRTSKKCAVIV--NKKRNSLRTSTKX 1196
 QY 220 QGPVQSSTDDIENLEGKATQNTENPREAGKPYECRVDVDPALYCTCRQPHNRRM 279

DB 1197 QADITSSSSSDIEDDDDQNSIEGSSDE---QKIPVTENLVLSHTGFC----- 1242
 QY 280 ICCRCHEWPHGDCVGVSEANGRL-----ERNEDYICPCTTILQVQ----DETN 326
 DB 1243 -----QSSGD-----EALSKSVPTVDDDDNDNENIACKMLLEBKIANISSBD 1289
 QY 327 GSATNEODSGCRSGVADGDTGCTSGIVGKSGEDGICGRLEKAMPKSKKLKIPQPV 386
 DB 1290 GSSDDEPFGKCRKTKQN-----EENQDEBA-KQVNSBSDDSDSBSKX----- 1333
 QY 387 EAPGAPKICIGKSSVAPDPSVCSNDILKHAATYFLLSGKEQKTPK--KVKTKP 444
 DB 1334 -----PR-----YRHLRH-KLTVDSGSGSEKTKPKKHKREKVGKN 1370
 QY 445 EKPSLPKSVGVGKISVHRLASKEKRNPKVMLASRBR-----TSGKRAACSSSTP 499
 DB 1371 RR-----KVSSDSB-DSDPQSGSVSEBSBDBQRPRTSAKKALEENOR 1417
 QY 500 SW-----ASDHVYNAVKPEKPEK 517
 DB 1418 SYKQKKRRRIKVGEDSSSEKNSSEEBERK 1449
 RESULT 8
 ATRX_PANTR STANDARD; PRT; 2492 AA.
 AC Q7YQW4;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
 nuclear protein) (XNP).
 GN ATRX.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Mammalia; Butheraia; Primates; Catarrhini; Homidae; Pan.
 NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22763540; PubMed=12777533;
 RA Kitano T., Schwarz C., Nickel B., Paabo S.;
 RT "Gene diversity patterns at 10 X-chromosomal loci in humans and
 chimpanzees";
 RL Mol. Biol. Evol. 20:1281-1289(2003).
 CC -1- FUNCTION: Could be a global transcriptional regulator. Modifies
 gene expression by affecting chromatin. May be involved in brain
 development and facial morphogenesis.
 CC -1- SUBUNIT: Probably binds E2H2. Binds annexin V in a calcium and
 phosphatidylcholine/phosphatidylethanolamine-dependent manner (By
 similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
 heterochromatin during interphase and mitosis, probably by
 interacting with HPI.
 CC -1- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC This SNF2-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL; AB102642; BAC8111.1; .
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR000330; SNF2_N.
 DR Pfam; PF00271; Helicase_C_1.
 DR Pfam; PF00176; SNF2_N_1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELICC; 1.

```

OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Heart, and Skeletal muscle;
RX MEDLINE=99428545; PubMed=10497235;
RA Kobayashi Y.M., Jones L.R.;
RT Identification of triadin 1 as the predominant triadin isoform
  expressed in mammalian myocardium.*
RL J. Biol. Chem. 274:28660-28668(1999).
CC -1- FUNCTION: May be involved in anchoring caldesmon to the
  junctional sarcoplasmic reticulum and allowing its functional
  coupling with the ryanodine receptor (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
  reticulum.
CC -1- ALTERNATIVE PRODUCTS:
  Event=Alternative splicing; Named isoforms=3;
  Name=Skeletal;
  IsoId=P82179-1; Sequence=Displayed;
  Name=Cardiac 1;
  IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
  Name=Cardiac 3;
  IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
CC -1- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF165916; AAF00222.1; -.
DR EMBL; AF165915; AAF00221.1; -.
DR EMBL; AF165917; AAF00223.1; -.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW Alternative splicing.
FT INIT MET 0
FT DOMAIN 1 46
FT TRANSMEM 47 67
FT DOMAIN 68 700
FT DOMAIN 74 74
FT CARBOHYD 616 616
FT CARBOHYD 257 277
FT VARSPLIC 278 700
FT VARSPLIC 466 466
FT VARSPLIC 524 579
FT VARSPLIC 700 AA; 78152 MW; F033E3A1BEE0C56 CRC64;
SQ
  Query Match 5.28; Score 169; DB 1; Length 700;
  Best Local Similarity 19.98; Pred. No. 0.034;
  Matches 114; Conservative 84; Mismatches 190; Indels 184; Gaps 24;
  3 DKGHS-----NEBAPKAIKPTSPKFEKRTWGRFRITIAKSGAGDTPEADPSEQ 52
  129 DKGIETEPPLAKQKRIHKKAKKKEKPERKILAK-----VAHKKKKKKKKKSEKK- 179
  53 QQHNIILSRSGROPKRTERVAEPLTVRRRGKQNVPLSDSEPTSSVTVDV----- 105
  180 AITHEKIKKKKBPETKMAKEBRKAKTBEIKIKGVKGQGEKVPITAKVBEVQKTPPKA 239

```

```

OY 106 -----ETASGVSSESS-----SEIRSG-----PVDSLGKHPASSEK 138
DB 240 KEKEGKETAFAVAKHQKQOYAFRCRYMIDMFVHGDIARPQSPALPPPLPTVQASRPTPASP 299
OY 139 AKGGBEBSDTSDSDGLTLKELONRLRRKREOQPERVSLRSGQNRLRKREBEDSAFTG 198
DB 300 TLBGRBEBEKKAQF-----KVTSE-YYKKEKEDV-----XKSDKOTAI-- 338
OY 199 SVQJGSAQDRPLCKQBPASQGP---VSQSETDIDENLEGKATQANTENPREAKPK 255
DB 339 -----DVEKGP--GKAPETKQGTIKVVAQAAKQDEKEDSKTKTPVEBHPK--GKKQ 389
OY 256 PECETVDPALALYICRGPHNNRPMICCRXCEMFHDCVGISEARGLLERGDEOYICN 315
DB 390 EKREKTY-----VEPAKSSKKEHSAPESEKQYK 415
OY 316 CTIIQVDETNGSATNEQSGCRSVAGADGTCTSIGTVEQK--SGEDQGIKGRIRKAAFPS 374
DB 416 AKTERAKETSAASTKKAIVGKK-----BEKTTTVQGEIRKESGTSSTASQKQEPB 468
OY 375 GKXKLIKIPQVVBAPGAPFCIGGCSVAOPDSVYCSNDCILKHAATNRPILSSGKEQXT 434
DB 469 IKXDEKM-----PKADKVRP-----KPPQS-----QVKKBEKS 497
OY 435 KKKKVKTKPEKPSLPKCSVQVQIKISVHKLASKREKREPVKKVWLASRSETSGKBAAC 494
DB 498 ESQVKEKAPQEQ-----DIKPEKTVSHQKPEBKVKQVKAT-----EKAAI 539
OY 495 ESST---PSMADHNNY---AVKPEKPEKPTA 520
DB 540 EKVTKPKPAKKAHQEKESPTIKTKDP-KPTS 570

```

RESULT 10
LAF4_HUMAN STANDARD; PRT; 1227 AA.

```

AC 01-OCT-1996 (Rel. 34, Created)
DB 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE LAF-4 protein (Lymphoid nuclear protein related to AP4).
GN LAF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96141096; PubMed=8555498;
RA Ma C., Staudt U.M.;
RT "LAF-4 encodes a lymphoid nuclear protein with transactivation
RT potential that is homologous to AP-4, the gene fused to MLL in
RT t(4;11) leukemias."
RL Blood 87:734-745(1996).
CC - FUNCTION: Putative transcription activator that may function in
CC lymphoid development and oncogenesis. Binds, in vitro, to
CC double-stranded DNA.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - TISSUE SPECIFICITY: Preferentially expressed in lymphoid tissues,
CC highest levels being found in the thymus.
CC - SIMILARITY: belongs to the AP4 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U34360; AAA98763.1; --
CC TRANSFAC: T04866; --
CC Genew; HGNC:6473; LAF4.

```

```

DR MIM: 601464; --
DR GO: GO:0005634; C:nucleus; TAS.
DR GO: GO:0007275; P:development; TAS.
DR InterPro: IPR007797; AP-4.
DR Pfam: PF05110; AP-4; 1.
KW Nuclear protein; Transcription regulation; Activator; DNA-binding.
FT DOMAIN 413 419 POLY-SER.
FT DOMAIN 422 432 POLY-SER.
FT DOMAIN 440 445 POLY-SER.
FT DOMAIN 670 679 POLY-SER.
SQ SEQUENCE 1227 AA; 133734 MW; 634B896FD7B9BBE7 CRC64;

```

Query Match 5.2%; Score 169; DB 1; Length 1227;
Best local similarity 19.1%; Pred. No. 0.063;
Matches 144; Conservative 95; Mismatches 209; Indels 306; Gaps 34;

```

OY 13 PRAIKETSEKFRKTGFRRTTAKREAGADTEADPSDQPOQHNLSKRS---GRQPKRT 69
DB 133 PAAV--PVQOSKGTGMQWK-----AGHPSPDGOQRATQOCSLFTLLGQVGRQOPRA 183
OY 70 ER---VEERFLTYRR-----RQKNVPVSL-----SEIRSG----- 91
DB 184 KQVCNVEVGLQTERPPAAMAHSSSGHCVQNPFSLASKPSLVQOKPTAYVRPMDQDQ 243
OY 92 --EDSGEPTSTVTDVETASGVSSESS-----SEIRSG----- 122
DB 244 APDESFKLSSSTSVHCTSYRGVPAKBPAPAKAKLSFISIPKQGBESGSETNSCVE 303
OY 123 -----PVS--DSLCKEHP-----ASSEKAKGEBEDTSDSDGL 156
DB 304 EIRREWTMLPPLSAIQAPKVEPTKPPFNKQSQLVSSGNNPKKQDABESBDNGTSNT 363
OY 157 TLKELONRLRRKREOQPERVSLRSGQNRLRKREBEDSAFTSGVOI--GSAQDRPLCKQ 214
DB 364 SM--LEDDIKLSDEE-----ENQQAQRTALRALSDSAVVOQPPKCR 405
OY 215 EPEASQAPVSQSETDIDENLEGKATQANTEE---NPREAKP---KPECEVYDPNA 265
DB 406 SVPSKSGSSSSSSSGTSSSSSDSESSGSDSEFTRESSSESGKPPHFSPEAB----- 459
OY 266 LYCICRPHNNRPMICCRCEWFHDCVGISBARGRLERNEEDYICPNCTIIQVODET 325
DB 460 -----PASNNKQWL-----DKWL-----NKVNPKPPILLQNES 488
OY 326 NSGATNEQSGCRSVAGADGTCTSIGTVEQKSGEDQGIKGRIRKAAFPSKKKLIKIPQV 385
DB 489 HGSBSNOY---YNPKEDVQDCKVDPVQCPSLRKRIKSTCKEBOQRPRTANK----- 538
OY 386 VEAPGA-----PKCIGPCSSVAOPDSVYCSNDCILKHAATNRPILSSGKE---Q 432
DB 539 --APSKRGVQKSPPAVAVVAAGAAAPPAVPCAP---AENADAPAR--RAGKKPRTARTE 592
OY 433 KTK-----PKKVKTKTP-----EKFS- 448
DB 593 RTSGAGDANCHREBPAAADALGTSVVVPPEPKTRCCNNRASHREKRLSSVTCRRRT 652
OY 449 -----LPR-----CSVQVGIKISSVHKRLASREKREPVKKVWLASRSETSG---KBA 493
DB 653 KGLSRIVPKSKERTERTSSSSSSSDLESBEQETPLSAQVVAASAGNQRLKEAA 712
OY 494 CESSTPSMASDHNNYNA-----VPEKPEKPTALSP----- 523
DB 713 ANGSQGPRAVGSINNRATTSDIANKLEBQVYTLVPFCRNELSLPLKSDSIRSLMWKIDL 772
OY 524 TLISKCTTH-PRAGPRC-----PSH 542
DB 773 TLISRIPEHLPOR--RGVLSAPATKQSBAPSH 804

```

RESULT 11
CGBP MOUSE STANDARD; PRT; 660 AA.
ID CGBP MOUSE
AC Q9CMT7;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB Cpg binding protein (Protein containing PHD finger and CXXC domain 1).
 GN CGBP OR PCCL1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,
 RA Aizawa K., Iwata M., Mishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanakawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Plechmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirnl L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guncionich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Borja A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschütter S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 RL -1- PUNCTON: Transcriptional activator that exhibits a unique DNA
 binding specificity for [AC]CpG[AC] unmethylated CpG motifs (By
 similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -1- SIMILARITY: Contains 1 CXXC-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.itsb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AK010337; BAB26862.1; -
 DR MGI: MGI:1921572; Cgdp.
 DR GO: GO:0016363; C:nuclear matrix; IDA.
 DR InterPro: IPR002857; Znf_CXXC.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF02008; ZF-CXXC; 1.
 DR SMART: SMO0249; PHD; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS0016; ZF_PHD_2; 1.
 DR Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
 KW Metal-binding; Coiled coil; Nuclear protein.
 KM ZN FING 28
 FT ZN FING 165 212 CXXC-TYPE.
 FT ZN FING 260 321 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 325 363 ARG/LYS-RICH (BASIC).
 FT DOMAIN 426 479 COILED COIL (POTENTIAL).
 FT DOMAIN 156 161 POLY-GLN.
 FT FT BIND 225 BY SIMILARITY.
 SQ SEQUENCE 660 AA; 76166 MW; 9D92CB6DA82069EA CRC64;

Query Match 5.2%; Score 167.5; DB 1; Length 660;
 Best local Similarity 18.4%; Pred. NO. 0.036; 72; Indels 237; Gaps 13;
 Matches 76; Conservative 29; Mismatches

QY 235 LRGATGANTTEENRBAKRPKPCCEVVDNALYCIQROPHNNRPMICDRGEWPHGDCV 294
 DB 1 MEG---DGSIDLEPPADGDSYSENGENAP--IYLCIRPDIINCFMGDCNMEVHGDCI 55
 QY 295 GISFARGTLERNRNGDYICPNC---TIIQVQ-----DETNGSATNEODSGCRS 339
 DB 56 RITERMAKAI---REMYCRECRERKDPRLTIRYHKKCRERDGSRRASBPBEGGRKR 111
 QY 340 VGAD-----GT----- 345
 DB 112 PASDPELQRRASGTGVGAMLRGSAAPKSSPQPLVATPSQHNNHQOQOQOQIIRSR 171
 QY 346 -----DC----- 353
 DB 172 MCGECACRTEDCGHPCFCDMKKFGSPNNKIRQKRLRQCLRRARBSYKTPSSLSPT 231
 QY 354 -----EOKSG---EDQG-----IKGRIERAMP----- 373
 DB 232 PSEALPRRRRPPYQQQPPQOSQKLRIRNDEGTVLSVVKPPFATNTPERLSDDLALD 291
 QY 374 -----SGKKKLKIFQ- 383
 DB 292 PDLVDPGAGAFDDHGLPMSDABSPFLDPLRRRAVKVKKRKKSEKKERRYKR 351
 QY 384 -----FVBAAPGPKICIGRCSSVAQPDSTYGSNDCLIKRAA 420
 DB 352 HRQKHQKDKWKKHPRERADAKDPAFLPQCLGRCVRAAGPSGRKCSDDCGKLAA 405
 RESULT 12
 CGBP HUMAN STANDARD; PRT; 656 AA.
 AC Q9P0U4; Q96BC8; Q9P2V7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB Cpg binding protein (Protein containing PHD finger and CXXC domain 1).
 GN CGBP OR PCCL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20153771; PubMed=10688657;
 RA Shin Voo K., Carlone D.L., Jacobsen B.M., Flodin A., Skalniak D.G.;
 RT "Cloning of a mammalian transcriptional activator that binds
 unmethylated CpG motifs and shares a CXXC domain with DNA
 methyltransferase, human trithorax, and methyl-CpG binding domain
 protein 1.";
 RT Mol. Cell. Biol. 20:2108-2121(2000).
 RN [2]
 RP SEQUENCE FROM N.A., AND DNA-BINDING.
 RX MEDLINE=20261256; PubMed=10799292;
 RA Fujino T., Hasegawa M., Shibata S., Kishimoto T., Imai S.-I.,
 RA Takano T.;
 RT "PCCL1, a novel DNA-binding protein with PHD finger and CXXC domain,
 is regulated by proteolysis.";
 RT Biochem. Biophys. Res. Commun. 271:305-310(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glaszi S.,
 RA Ansgore W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
 RA Tauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
 RA Mewes H.-W., Oesterwalder B., Obermaier B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs.";
 RL Genome Res. 11:422-435(2001).
 RN [4]
 RP SEQUENCE FROM N.A.

RC TISSUE-Colon, and Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.C.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenn C.M., Schuler G.D.,
 RA Altschul S.F., Zeefer B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Datchenko L., Maruska K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stopleton M., Soares M.B., Bonaldi M.F., Caesavari T.L., Scheetz T.B.,
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loughellano N.A., Peters G.J., Abramson R.D., Millaby S.J.,
 RA Boeck S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton B., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.J., Skalek U., Smallus D.E.,
 RA Scherren A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP DNA-BINDING DOMAIN.
 RX MEDLINE=21576202; PubMed=11572867;
 RA Lee J.H., Voo K.S., Skalniak D.G.,
 RT "Identification and characterization of the DNA binding domain of
 RT Cpg-binding protein.";
 RL J. Biol. Chem. 276:44669-44676(2001).
 CC -1- FUNCTION: Transcriptional activator that exhibits a unique DNA
 CC binding specificity for [AC]CpG[AC] unmethylated Cpg motifs.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- DOMAIN: The acidic domain carries the potential to activate
 CC transcription.
 CC -1- PTM: May be regulated by proteolysis.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -1- SIMILARITY: Contains 1 CXXC-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, APL14758; AAP37799.1; -
 CC EMBL, AB031069; BAA96307.1; -
 CC EMBL, AL136862; CAB66796.1; -
 CC EMBL, BC014940; AAH14940.1; -
 CC EMBL, BC015733; AAH15733.1; -
 CC GO: GO:0005634; C:nucleus; IDA.
 CC GO: GO:0016563; P:transcriptional activator activity; IDA.
 CC GO: GO:0045322; P:unmethylated Cpg binding; IDA.
 CC GO: GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
 CC InterPro: IPR002857; Znf_CXXC.
 CC Pfam: PF00628; PHD; 1.
 CC Pfam: PF00628; PHD; 1.
 CC SMART: SM00249; PHD; 1.
 CC PROSITE: PS01359; ZF_PHD_1; 1.
 CC PROSITE: PS50016; ZF_PHD_2; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
 CC Metal-binding; Coiled coil; Nuclear protein.
 CC -----
 CC ZN FING 28 76
 CC ZN FING 161 208
 CC ZN FING 326 317
 CC ZN FING 321 359
 CC ZN FING 422 474
 CC ZN FING 156 161
 CC ZN FING 162 221
 CC ZN FING 117 117
 CC D -> N (IN RBP. 1).

PT CONFLICT 302 302 H -> N (IN RBP. 1).
 FT CONFLICT 340 340 K -> KWMER (IN REF. 4; AAH15733).
 SQ SEQUENCE 656 AA; 75711 MW; 6D2376B44905A18 CRC64;
 Query Match 5.2%; Score 166.5; DB 1; Length 656;
 Best Local Similarity 18.4%; Pred. No. 0.043;
 Matches 74; Conservative 31; Mismatches 55; Indels 243; Gaps 13;
 QY 245 BENPBAKPKKPCRYVDPNALYTCROPHANRPIMCCRCSEMPFGDVGISEAGRL 304
 DB 15 BDKSENGENAP-----IYCIKRPDINCFMIGCDNCNEMFAGDCIRITEKNAKAL 65
 QY 305 BENGEDYICPNC-----TILVO-----DETNGSATNEODS----- 335
 DB 66 -----REWYCRBRKEDPKLEIRYKHKSKERDQ-----NEDSSERPDGGRKRPVDDPD 118
 QY 336 -----GCRSVGADGT----- 345
 DB 119 QRAGSGTGVGAMLRGASPHKSSPQPLVATPSPGHQOQOQOIKRSAMGCEACRR 178
 QY 346 -DC----- 353
 DB 179 BDCGHCDFCRDMKKKGGPNKIRQKCLRQOQDRARESYRPPSSLPVPPSSLP 238
 QY 354 -----EOKSG-----BDQ-----IKRIEKANP----- 373
 DB 239 LPTQOQPOPSQKLGRIREDGAVASSTVKEPPEATPTPLDDEDLPLPDLYQDFCAGA 296
 QY 374 -----SGKKKLKIPQ----- 383
 DB 299 PDHGLPMSDTEBSPFLDPALRKAVKGVYRREKSEKKEBYKRRQKHKDKW 358
 QY 384 -----PVTEAPGAPKCTIPGCCSVAPQPSVYCSNDILKHA 420
 DB 359 KHPERADADPASPPLPQCLGPGCVRAPOPSKTYCSDCGMKLAA 401
 DB
 RESULT 13
 DMP1_RAT
 ID DMP1_RAT STANDARD; PRT; 489 AA.
 AC P98193;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
 DE protein-1) (DMP-1) (AG1).
 GN DMP1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Tooth;
 RX MEDLINE=93286101; PubMed=8509401;
 RA George A., Sabay B., Simonian P.A., Veis A.;
 RA "Characterization of a novel dentin matrix acidic phosphoprotein.
 RA Implications for induction of biomineralization.";
 RT J. Biol. Chem. 268:12624-12630(1993).
 RL -1- FUNCTION: Could be involved in the induction of mineralization of
 RL extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Expressed in tooth particularly in odontoblast
 CC and ameloblast.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, L11354; -; NOT_ANNOTATED_CDS.

Biomineralization; Extracellular matrix; Signal.

FT SIGNAL 1 16 POTENTIAL.

FT CHAIN 17 489 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.

FT DOMAIN 41 44 POLY-PRO.

FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SITE 334 336 CELL ATTACHMENT SITE (POTENTIAL).

SQ SEQUENCE 489 AA; 53058 MW; 59P8361479DDA085 CRC64;

Query Match 5.1%; Score 165.5; DB 1; Length 489;

Best Local Similarity 26.0%; Pred. No. 0.035;

Matches 72; Conservative 45; Mismatches 97; Indels 63; Gaps 12;

QY 3 DKGF-----LSNEAPKAIKPTSK-----EPRKFWKRTTAKBAGADTEAD 46

DB 205 DRGHTRMSSAGIRSEBSKGDHEPTSTQSDSDSQVEFSGRSKFRSRVSEEDRGEL-AD 263

QY 47 PSEQQPOQHNLSTARSGQPKRTVERVEFLTVRRKGNKPNVSLDESEPTSTVTDVE 106

DB 264 SNSRSTQSVSTEDFRS-KESRSSTQED--TAETQSQEDSPEDPSSSS-----E 312

QY 107 TARGSVSSSRIRSGPVSDSLGKEHPASSR-----KAKGGBEE-----EDT 148

DB 313 BAGEPSQSSSESGGVASERGNPDNTSGTGQORDESSESDPLNTFSSSSQSTEEQ 372

QY 149 SDSIS-DGLTLKEIKNLRRKREQEPVRSLSGQNLRLKK--RRRSDAFTGVSQVIGSA 205

DB 373 GDSSNSLSLSSE-----ESGSAQDEDSGGLQSGASRSRSQESQSDRS 424

QY 206 EQRPLKQ-----EPASQPVSSQSTDDIENOLE 236

DB 425 EENRDSQDSRSRKRBSNSTGTSSSEDDHPKNIE 461

RESULT 14

AK12_HUMAN STANDARD; PRT; 1781 AA.

AC Q02952; Q00310; Q00498; Q99970;

DT 01-FEB-1994 (Rel. 28, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE A-kinase anchor protein 12 (A-kinase anchor protein 250 kDa) (AKAP 250) (Myoasthenia gravis autoantigen gravin).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A. (ISOFORM 1).

RP TISSUE=Heart;

RC MEDLINE=97153077; PubMed=9000000;

RA Nauert J.B., Klauk T.M., Langeberg L.K., Scott J.D.;

RT "Gravin, an autoantigen recognized by serum from myoasthenia gravis patients, is a kinase scaffold protein.";

RL Curr. Biol. 7:52-62(1997).

[2]

SEQUENCE FROM N.A. (ISOFORM 2).

RP TISSUE=Umbilical vein endothelial cells;

RC MEDLINE=98269042; PubMed=9604001;

RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;

RT "Changes of gene expression by lysophosphatidylcholine in vascular endothelial cells: 12 up-regulated distinct genes including 5 cell growth-related, 3 thrombosis-related, and 4 others.";

RL J. Biochem. 123:1119-1126(1998).

[3]

SEQUENCE OF 43-1781 FROM N.A.

RP TISSUE=Umbilical vein endothelial cells;

RA Bowditch R.D., Ginsberg M.H.;

RT "Sequence of gravin cDNA isolated from a human umbilical vein endothelial cell library.";

RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.

[4]

RP SEQUENCE OF 1477-1781 FROM N.A.

RC TISSUE=Umbilical vein endothelial cells;

RX MEDLINE=92395179; PubMed=1522245;

RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R., Lindstrom J., Ginsberg M.H.;

RT "Molecular cloning and preliminary characterization of a novel cytoplasmic antigen recognized by myasthenia gravis sera.";

RL J. Clin. Invest. 90:992-999(1992).

CC -1 FUNCTION: Anchoring protein kinase A (PKA) and protein kinase C (PKC).

CC -1 SUBUNIT: Binds to dimeric RII-alpha regulatory subunit of PKC.

CC -1 SUBCELLULAR LOCATION: Cytoplasmic. May be part of the cortical cyoskeleton.

CC -1 ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q02952-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q02952-2; Sequence=VSP_004110, VSP_004111;

CC Note=No experimental confirmation available;

CC -1 TISSUE SPECIFICITY: Expressed in endothelial cells, cultured fibroblasts and osteosarcoma, but not in platelets, leukocytes, monocytic cell lines or peripheral blood cells.

CC -1 INDUCTION: Activated by lysophosphatidylcholine (lysoPC).

CC -1 DOMAIN: Polypasic regions located between residues 265 and 556 are involved in binding PKC.

CC -1 DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY PATIENTS WITH MYASTHENIA GRAVIS (MG).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL; U01607; AAC51366.1; -

DR EMBL; AF001504; AAB58938.1; -

DR EMBL; AB003476; BAA19927.1; -

DR EMBL; M96322; AAA35911.1; -

DR GeneW; HGNC:370; AKAP12.

DR MIM; 604698; -

DR GO; GO:0005737; C:cytoplasm; TAS.

DR GO; GO:0005079; F:protein kinase A anchoring activity; TAS.

DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.

DR InterPro; IPR001573; PKtna anch.

DR Pfam; PF03832; PKtna anch; 3.

DR Antigen; Repeat; Alternative splicing.

KW DOMAIN 603 633 AKAP 1.

FT DOMAIN 752 782 AKAP 2.

FT DOMAIN 797 827 AKAP 3.

FT DOMAIN 98 101 POLY-GLU.

FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).

FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).

FT VARSPLIC 1 98 Missing (in isoform 2).

FT VARSPLIC 1 98 /FTId=VSP_004110.

FT VARSPLIC 99 106 REEIVTR -> MGLTITIT (in isoform 2).

FT /FTId=VSP_004111.

FT CONFLICT 117 117 E -> K (IN REF. 2).

FT CONFLICT 142 144 NEN -> TPEI (IN REF. 2 AND 3).

FT CONFLICT 215 215 Q -> K (IN REF. 2).

FT CONFLICT 448 448 G -> E (IN REF. 2 AND 3).

FT CONFLICT 694 694 R -> G (IN REF. 2 AND 3).

FT CONFLICT 867 867 G -> S (IN REF. 2 AND 3).

FT CONFLICT 986 986 S -> A (IN REF. 2 AND 3).

FT CONFLICT 1530 1530 E -> E (IN REF. 3 AND 4).

FT CONFLICT 1581 1581 V -> M (IN REF. 4).

FT CONFLICT 1601 1601 Q -> L (IN REF. 2).

SQ SEQUENCE 1781 AA; 191439 MW; BAA13937379PACOF CRC64;

Query Match 5.1%; Score 165.5; DB 1; Length 1781;
 Best Local Similarity 20.4%; Pred. No. 0.15;
 Matches 133; Conservative 81; Mismatches 201; Indels 237; Gaps 32;

```

QY 3 DKGLTSLN---EAPKA-----IKPTSKE-----FRKTW---GFRRT-----32
DB 136 DDGGEENRNIETQSSSNLELQTPTESQANDIGFKVFYFGVFKFVKDKTEKPTV 195
QY 33 ---TIAREGAGTBA---DPS-----EQOPOHNLSTARSGRPKRTREVERFL 76
DB 196 QLLTVKXDBEGGAGADHPDPSLGAAPAAKSESPKQ-----STKPEPTL 242
QY 77 TTVARRKQAVPVLSDSSSEPTSSVTVDVETASGVSSESRSPVSDSLGK-----130
DB 243 KRBSHABISPPA---ESGQVBERCKEGERKQEKPEKSKAESPTSETSGTFKKEFP 299
QY 131 -----EHPASSKAKGEEBEDTSDSDGLTLKELQNLRRKREQPVERSLRGSQN 183
DB 300 TQGMAGMRKKTSPFKPK---EDEVEASB-----KKKEQGP-----331
QY 184 RLARKRRED-SATSGVQISAGQDRPLCKQEP-BASQGVSGSETDDI-----ENQLEG 237
DB 332 --EKVTEBBDKAAVASAKLTASQAHF---QEPASAHBRLSAEYKVELPSEBQVSG 386
QY 238 KATQNTBENPRBAGKPKP--ECVYDPVALYICIGQPHNRFMICCDRCBEMFHQDCVGI 296
DB 387 --SGSPSEB-----KPAPLATEVFDKTL-----EVQEEVVA-----EVHV 420
QY 297 SEARGRLLENGBDYICPNCITLQVQDETNGSATNEODSGRGVAGDGTCTSIGTVQK 356
DB 421 STVERTEBQKTE-----VEBTAGSVPAELVGMADAEQBARPAKELKLTKE 468
QY 357 --SGED--QGI-----KGRLEKANPSGKKKLT-----PPVVEAPAPAKPI 395
DB 469 CVSGEDPTQGADLSPDEKYLKSPREGVAVSEVEMLSQGRMVQGSPLKLTFTSTGLKLS 528
QY 396 G-----PGCSVAPQPSVYCSNDCLIKHAAA 421
DB 529 GKQKQKRGCGDESGEHTQVPADSPSOEBOKGSSASSPPEPRLTC-----LEKLA 583
QY 422 TMRPLSGKEQKTYKPEKVK-----TYKEKPSLKCSVQVGIKISSVHKLASKEKNP 475
DB 584 EVQDDGEABEGATSGEKKRGVTPWASFKQVTPKKEV-----RRPSEBDEKDB 633
QY 476 VKYVWLASRSETSGKAAACESTPESWASDHNTNAYKPEKPTLSPTLIS 527
DB 634 LDYKVSATLSST-----BSTASEMQEBMKGSVEBPKPEPKKVDTSVS 677

```

RESULT 15
 UN89 CAEEL STANDARD; PRT; 6632 AA.

AC 001761; Q17362;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
 GN UNC-89 OR C09D1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodertinae; Caenorhabditis.
 OX NCB1_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN=Bristol N2;
 RX MEDLINE=96180278; Pubmed=8603916;
 RA Benlan G.M., Tinley T.L., Tang X., Borodovsky M.;
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
 RT assembly, encodes a giant modular protein composed of Ig and signal
 RT transduction domains";
 RL J. Cell Biol. 132:835-848 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;
 RA Du Z., Le T.T., Wilson R.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Structural component of the muscle M-line. Myofibrillar
 CC lattice assembly begins with positional cues laid down in the
 CC basement membrane and muscle cell membrane. UNC-89 responds to
 CC these signals, localizes, and then participates in assembling an
 CC M-line.
 CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
 CC -!- SIMILARITY: Contains 1 DBP-homology (DB) domain.
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 1 PH domain.
 CC -!- SIMILARITY: Contains 5 RGD domains.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sib-ch).

CC EMBL: U33058; AB00542.1; --
 CC EMBL: AF003131; AB054132.2; --
 CC PDB: 1PHO; 20-DEC-00.
 DR WormRep; C09D1.1; CE30426.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR007850; RGD.
 DR InterPro; IPR000219; RGDGF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF00047; Ig; 47.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF05177; RGD; 5.
 DR Pfam; PF00621; RGDGF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00408; IGC2; 23.
 DR SMART; SM00325; RGDGF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50010; DR 2; 1.
 DR PROSITE; PS50835; IG-LIKE; 49.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 1.
 KM Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
 KM 3D-structure.
 FT DOMAIN 63 127 SH3.
 FT DOMAIN 152 330 PH.
 FT DOMAIN 342 498 PH.
 FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
 FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
 FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
 FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
 FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
 FT DOMAIN 1272 1315 THR-RICH.
 FT DOMAIN 1375 1475 RGD 1.
 FT DOMAIN 1479 1585 RGD 2.
 FT DOMAIN 1597 1695 RGD 3.
 FT DOMAIN 1700 1799 RGD 4.
 FT DOMAIN 1800 1860 RGD 5.
 FT DOMAIN 1862 2067 IG-LIKE C2-TYPE 7.
 FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:19:55 ; Search time 48 Seconds
(without alignments)
3526.319 Million cell updates/sec

Title: US-09-787-016a-4
Perfect score: 3228
Sequence: 1 MDDKGLHSEBAPKAIKETS.....LKSQVGLMKAISYSPFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1124875 seqs, 275673149 residues
Total number of hits satisfying chosen parameters: 1124875

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	740	22.9	181	9	US-09-925-301-1169 Sequence 1169, App
2	194.5	6.0	676	10	US-09-823-187-24 Sequence 24, App
3	194.5	6.0	676	10	US-09-863-776-14 Sequence 14, App
4	191.5	5.9	676	14	US-10-203-708-28 Sequence 28, App
5	191.5	5.9	717	14	US-10-203-708-29 Sequence 29, App
6	190.5	5.9	518	9	US-09-925-299-862 Sequence 862, App
7	190.5	5.9	518	10	US-09-925-299-862 Sequence 862, App
8	190.5	5.9	667	10	US-09-863-776-16 Sequence 16, App
9	190.5	5.9	667	10	US-09-863-776-18 Sequence 18, App
10	188.5	5.8	555	14	US-10-106-698-5847 Sequence 5847, App
11	188.5	5.8	670	10	US-09-823-187-86 Sequence 86, App
12	188.5	5.8	670	10	US-09-863-776-51 Sequence 51, App
13	188.5	5.8	671	10	US-09-946-374-308 Sequence 308, App
14	188.5	5.8	671	10	US-09-823-187-85 Sequence 85, App
15	188.5	5.8	671	10	US-09-863-776-20 Sequence 20, App

16	188.5	5.8	671	12	US-10-206-915-346 Sequence 346, App
17	188.5	5.8	671	12	US-10-199-670-346 Sequence 346, App
18	188.5	5.8	671	12	US-10-201-858-346 Sequence 346, App
19	188.5	5.8	671	12	US-10-205-890-346 Sequence 346, App
20	188.5	5.8	671	12	US-10-208-024-346 Sequence 346, App
21	188.5	5.8	671	12	US-10-201-853-346 Sequence 346, App
22	188.5	5.8	671	12	US-10-174-581-346 Sequence 346, App
23	188.5	5.8	671	12	US-10-176-483-346 Sequence 346, App
24	188.5	5.8	671	12	US-10-176-749-346 Sequence 346, App
25	188.5	5.8	671	12	US-10-176-914-346 Sequence 346, App
26	188.5	5.8	671	12	US-10-176-915-346 Sequence 346, App
27	188.5	5.8	671	12	US-10-006-485A-308 Sequence 308, App
28	188.5	5.8	671	12	US-10-013-907A-308 Sequence 308, App
29	188.5	5.8	671	12	US-10-015-499A-308 Sequence 308, App
30	188.5	5.8	671	12	US-10-176-484-346 Sequence 346, App
31	188.5	5.8	671	12	US-10-180-550-346 Sequence 346, App
32	188.5	5.8	671	12	US-10-183-014-346 Sequence 346, App
33	188.5	5.8	671	12	US-10-187-728-346 Sequence 346, App
34	188.5	5.8	671	12	US-10-187-740-346 Sequence 346, App
35	188.5	5.8	671	12	US-10-187-883-346 Sequence 346, App
36	188.5	5.8	671	12	US-10-194-460-346 Sequence 346, App
37	188.5	5.8	671	12	US-10-194-463-346 Sequence 346, App
38	188.5	5.8	671	12	US-10-194-463-346 Sequence 346, App
39	188.5	5.8	671	12	US-10-194-484-346 Sequence 346, App
40	188.5	5.8	671	12	US-10-195-884-346 Sequence 346, App
41	188.5	5.8	671	12	US-10-195-896-346 Sequence 346, App
42	188.5	5.8	671	12	US-10-196-744-346 Sequence 346, App
43	188.5	5.8	671	12	US-10-196-755-346 Sequence 346, App
44	188.5	5.8	671	12	US-10-196-757-346 Sequence 346, App
45	188.5	5.8	671	12	US-10-197-704-346 Sequence 346, App

ALIGNMENTS

RESULT 1
US-09-925-301-1169
Sequence 1169, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P4106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1169
LENGTH: 181
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-301-1169

Query Match 22.9%; Score 740; DB 9; Length 181;
Best local similarity 85.1%; Pred. No. 3.2e+46;
Matches 131; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY 252 GKRPECEVYDPAALYICIGRPPNNRFMCCDRCEWPHQDVGISBARGRLIERNGEDY 311
DB 11 GRPKPCCEGYDPAALYICIGRPPNNRFMCCDRCEWPHQDVGISBARGRLIERNGEDY 70
QY 312 ICPNCTIIQVQDHTSHETADQSEAKRRPDADCTGTSIGTIRQSSBDGIGRIEKAA 371
DB 71 ICPNCTIIQVQDHTSHETADQSEAKRRPDADCTGTSIGTIRQSSBDGIGRIEKAA 130
QY 372 NPSGKKKLTIFQVYVAPGAPKICIGPGCSSVAQP 405
DB 131 NPSGKKKLTIFQVYVAPGAPKICIGPGCSSVAQP 164

```
RESULT 2
US-09-823-187-24
; Sequence 24, Application US/09823187
; Publication No. US20030096952A1
GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Liu, Xiaohong
; APPLICANT: Majumder, Kumud
; APPLICANT: Padigaru, Muralidhar
; APPLICANT: Patnajan, Meera
; APPLICANT: Shukets, Richard A
; APPLICANT: Spaderma, Steven K
; APPLICANT: Spytek, Kimberly
; APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; PIR REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/193,339
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/193,205
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/195,343
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: 60/195,088
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,005
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,792
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 60/196,556
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: 60/197,081
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/197,525
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/197,087
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-823-187-24

Query Match      6.0%; Score 194.5; DB 10; Length 676;
Best Local Similarity 22.0%; Pred. No. 1,6e-05;
Matches 118; Conservative 83; Mismatches 183; Indels 153; Gaps 24;

QY 28 GFRRTT-----IAKREGAGDTFAD-----PSEOCPOQNNLSLRSGRQPKTERVERFL 76
DB 154 GLKRTPLALKVSVKARARASDLDQASVSPSEENSSSSSEKTSDDPTPEK--KAA 211
QY 77 TTVRRRG-----KXNVPSLEDSSEPTSTVTVDVTAEGSVSSSEIRSGFVSLSLK 130
DB 212 VVAPRRGPGIKKKKAPASDSDSCADSDGAKPEVVAARASASSSSSSSSDSDSVYK 271
QY 131 EHPASSE-----KAKGGBEED-----TSSDSDEL-TLKELONR-----LRR 167
DB 272 KPPRGKPAKBPVPKRGGKPKPERPPSSSSSSSDSDSDDEVIRSEWRREARRRELEARR 331
QY 168 KKEQEPVERSLRGSQNRLLKRGREBDSATGVSQVISA-----EODRPLCKQPEA- 218
DB 332 RKEQBEELRLRKEQBEERERRRR--ADRGAEKRSQSSSGDELAEDEBVPVKKRKR 389
QY 219 SGGPVQSSETDDIENGLGKATQANTENPREAK--PPECEVYDPAALYCTICROPHNN 276
DB 390 GRGPSSSSBP-EABLE-----REAKKSAKPPQSSSTBP-----ARKP----- 427
QY 277 RRMICDRCBEMPHGDCVGISPAKGRLLERNGEDYICPCTIIQVQDFTNGSATINQDSG 336
```

```
DB 428 -----GQKKEVR-----PEERQ 441
QY 337 CRSYGADGTCTSIGTWBQKSGEODGIGRIEKAHPGKKTLL-----IFQPVVAPG 390
DB 442 ARPVVERT-----RKRSBGSMDRKVKKKKPEVEERLQQLHSEIKFALRVDSPD 492
QY 391 APKICGPGCSSVAPDPSVYCSNDCLKH--AAATWRFLSSGKQKTPREKVKTPKPKPS 448
DB 493 VRCL-----NALBEIGTQVTSQILQKNTDVATLK-----KIRRYKANKDVMEKQABVY 543
QY 449 LPKCSVQVGIKISVHK-RLASERKRPVKVMLASRSETSGKAAACSSSTPSWASD 504
DB 544 TRKSRVVGPKIEAVGYKNGMEKKAERKL---AGELAGEELAGEBAPEKALD 597

RESULT 3
US-09-863-776-14
; Sequence 14, Application US/09863776
; Publication No. US20030198953A1
GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhar
; APPLICANT: Spaderma, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Raateilil, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangolli, Bsha
TITLE OF INVENTION: No. US20030198953A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
CURRENT APPLICATION NUMBER: US/09/863,776
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206,679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207,798
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/208,263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208,831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209,451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210,060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219,507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221,337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221,927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263,135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,688
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263,694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
; LENGTH: 676
; TYPE: PRT
```

ORGANISM: Homo sapiens
US-09-863-776-14

Query Match 6.0%; Score 194.5; DB 10; Length 676;
Best Local Similarity 22.0%; Pred. No. 1.6e-05;
Matches 118; Conservative 83; Mismatches 183; Indels 153; Gaps 24;

```
QY 28 GFRRTT-----IARBGAGDTBAD-----PSEQOPOQHNLRLRSRGROPKTERVEEFL 76
DB 154 GLKRTTALMKSYSRRARASDLDQASVPSBERSESSSEKTSDDFTPEK--KAA 211
DB 77 TTVRRG-----KKNVPSLEDSSEPTSTVTDVETASBGVSSEBIRSGVSDSLGK 130
DB 212 VRAPRRGLGRRKKKAPASDSKADSDGAKPEPVANARASSSSSSSSDSDSVYK 271
QY 131 EHPASB-----KAKGEBEED-----TSDSDG--TLKEI--LNR 167
DB 272 KPRGRKPAKPLPKPRGRKPKPRPPSSSSSDSDSDVDRISEMKRDBARRRLEAAR 331
QY 168 KROBPVERSLRGQNRLLKKREEDSAGTSVQISG-----EODRPLCKQPEA- 218
DB 332 RROGBEELRLRLEQGEKEKERRER--ADRGARSGSGSSGDELREDDPEVKKRGKR 389
QY 219 SGQPVQSSTDDIENQLEKATQNTENPREAK--PRECEVDPNALYICROPHNN 276
DB 390 GRGPPSSSDSE-BAELR-----REAKSAKQPQSSSTP-----ARKP--- 427
QY 277 RFMICCDRCBWFHGDVGISEARGLLRNGEDYICPNCTILQVDNTGSAATNBQDSG 336
DB 428 -----GQKEKRV-----PEEKQ 441
QY 337 CSVGADGTDCTSIGTVEKSGEDGCIKRIEKAANPSGKKLK-----IFQPVYEA 390
DB 442 AARVAVERT-----KRSEGFMDRKVEKKEKPSVEKIQKLSHETKFLATXDSPD 492
QY 391 APRCIGPGCSVAQPDVYVCSNDCILKH--AAATMFLSGKEQKTKPKKVTKEPKFS 448
DB 493 VRCL-----NALELGTLOVTSQILQKNTDVATLK-----KIRRYANQDVMEKAAEVY 543
QY 449 LKCSYOVGKIKISSVHK--RLASEKENPVKKVWLARSSTSGEAACESSTPSWASD 504
DB 544 TLKSRVLPKTLBAVQKVNKAQMEKEKAHEKL--AGEBELAGBELAGEBAPOKEAED 597

RESULT 4
US-10-203-708-28
/ Sequence 28, Application US/10203708
/ Publication No. US20030149238A1
/ GENERAL INFORMATION:
/ APPLICANT: SMITHKLINE BEECHAM CORPORATION
/ APPLICANT: SMITHKLINE BEECHAM P.1.C.
/ TITLE OF INVENTION: NOVEL COMPOUNDS
/ FILE REFERENCE: GP50013
/ CURRENT APPLICATION NUMBER: US/10/203,708
/ CURRENT FILING DATE: 2002-08-13
/ PRIOR APPLICATION NUMBER: PCT/US01/04703
/ PRIOR FILING DATE: 2001-02-14
/ PRIOR APPLICATION NUMBER: 60/182,172
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: 60/186,084
/ PRIOR FILING DATE: 2000-02-29
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 28
/ LENGTH: 676
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-203-708-28
```

Query Match 5.9%; Score 191.5; DB 14; Length 676;
Best Local Similarity 21.8%; Pred. No. 2.6e-05;
Matches 117; Conservative 84; Mismatches 183; Indels 153; Gaps 24;

```
QY 28 GFRRTT-----IARBGAGDTBAD-----PSEQOPOQHNLRLRSRGROPKTERVEEFL 76
DB 154 GLKRTTALMKSYSRRARASDLDQASVPSBERSESSSEKTSDDFTPEK--KAA 211
DB 77 TTVRRG-----KKNVPSLEDSSEPTSTVTDVETASBGVSSEBIRSGVSDSLGK 130
DB 212 VRAPRRGLGRRKKKAPASDSKADSDGAKPEPVANARASSSSSSSSDSDSVYK 271
QY 131 EHPASB-----KAKGEBEED-----TSDSDG--TLKEI--LNR 167
DB 272 KPRGRKPAKPLPKPRGRKPKPRPPSSSSSDSDSDVDRISEMKRDBARRRLEAAR 331
QY 168 KROBPVERSLRGQNRLLKKREEDSAGTSVQISG-----EODRPLCKQPEA- 218
DB 332 RROGBEELRLRLEQGEKEKERRER--ADRGARSGSGSGDELREDDPEVKKRGKR 389
QY 219 SGQPVQSSTDDIENQLEKATQNTENPREAK--PRECEVDPNALYICROPHNN 276
DB 390 GRGPPSSSDSE-BAELR-----REAKSAKQPQSSSTP-----ARKP--- 427
QY 277 RFMICCDRCBWFHGDVGISEARGLLRNGEDYICPNCTILQVDNTGSAATNBQDSG 336
DB 428 -----GQKEKRV-----PEEKQ 441
QY 337 CSVGADGTDCTSIGTVEKSGEDGCIKRIEKAANPSGKKLK-----IFQPVYEA 390
DB 442 AARVAVERT-----KRSEGFMDRKVEKKEKPSVEKIQKLSHETKFLATXDSPD 492
QY 391 APRCIGPGCSVAQPDVYVCSNDCILKH--AAATMFLSGKEQKTKPKKVTKEPKFS 448
DB 493 VRCL-----NALELGTLOVTSQILQKNTDVATLK-----KIRRYANQDVMEKAAEVY 543
QY 449 LKCSYOVGKIKISSVHK--RLASEKENPVKKVWLARSSTSGEAACESSTPSWASD 504
DB 544 TLKSRVLPKTLBAVQKVNKAQMEKEKAHEKL--AGEBELAGBELAGEBAPOKEAED 597

RESULT 5
US-10-203-708-29
/ Sequence 29, Application US/10203708
/ Publication No. US20030149238A1
/ GENERAL INFORMATION:
/ APPLICANT: SMITHKLINE BEECHAM CORPORATION
/ APPLICANT: SMITHKLINE BEECHAM P.1.C.
/ TITLE OF INVENTION: NOVEL COMPOUNDS
/ FILE REFERENCE: GP50013
/ CURRENT APPLICATION NUMBER: US/10/203,708
/ CURRENT FILING DATE: 2002-08-13
/ PRIOR APPLICATION NUMBER: PCT/US01/04703
/ PRIOR FILING DATE: 2001-02-14
/ PRIOR APPLICATION NUMBER: 60/182,172
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: 60/186,084
/ PRIOR FILING DATE: 2000-02-29
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 29
/ LENGTH: 717
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-203-708-29
```

Query Match 5.9%; Score 191.5; DB 14; Length 717;
Best Local Similarity 21.8%; Pred. No. 2.8e-05;
Matches 117; Conservative 84; Mismatches 183; Indels 153; Gaps 24;

```
QY 28 GFRRTT-----IARBGAGDTBAD-----PSEQOPOQHNLRLRSRGROPKTERVEEFL 76
DB 195 GLKRTTALMKSYSRRARASDLDQASVPSBERSESSSEKTSDDFTPEK--KAA 252
DB 77 TTVRRG-----KKNVPSLEDSSEPTSTVTDVETASBGVSSEBIRSGVSDSLGK 130
DB 253 VRAPRRGLGRRKKKAPASDSKADSDGAKPEPVANARASSSSSSSSDSDSVYK 312
```



```

Qy 131 EHPASSB-----KAKGGBEED-----TSDSDGL-TLKEIQNR-----LRR 167
Db 313 KPRGRKPAKRLPFRGRKRPFRPSSSSSDSDSDVDRISBKRDEARRELEARR 372
Qy 168 KRGEPVRSIRGSONRLRKRRREDSABTGSVOIGSA-----BDRPLCKQBPXA- 218
Db 373 RRGGBEELRLRLEQKEKERRRRR--ADRGAEKSGSGSSGDBLEBDEPVRKGRGR 430
Qy 219 SGPVSGSETDDIENQLEKATQGTENPREAGK--PKPECEYDPMALYICIQPHNN 276
Db 431 GRGPSSSDSBR--BAELR-----REAKSACKPOSSSTEP-----AKRP--- 468
Qy 277 RPMICDRCEWPHDDCVGISRAGRLEBNGEDYICNPCTILQVQDETNGSATNEQDSG 336
Db 469 -----GQEKERYR-----PBEKQ 482
Qy 337 CRVYAGDGTCTSIGTVBOKSGBDGIGRIEKAAPSGKKLK-----IFQVYBAAG 390
Db 483 AKPVYVERT-----RKSGSGSMRKRYEKKGPVBERLOQLHSIRKALKVDSPD 533
Qy 391 APKCIGPCSSVAQPDVYCSNDCTIKR--AAATYRPLSSGEGQTKPKREKVTPEKFS 448
Db 534 VKRCL-----NMLEBGLQVTSQILQKNTDVVATLK-----KIRRYKANKDWMERKAAYV 584
Qy 449 LPKGVQVGIKISSYHK--RLASERENPVKVMLASRSETSGKEAACSSSTPSASD 504
Db 585 TRLSRVLPPIEAVQVYKAKMEKAKBEKL--AGEBLAGEBLAGEBAPOKAEAD 638

```

RESULT 6 US-09-925-299-862

```

; Sequence 862, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862

```

Query Match 5.9%; Score 190.5; DB 9; Length 518;

Best Local Similarity 22.0%; Pred. No. 2.2e-05;

Matches 119; Conservative 80; Mismatches 185; Indels 157; Gaps 25;

```

Qy 28 GFRRT-----IARBGAGTDEAD-----PSEQPOQHNLRLRSGRQPKTERVBEFL 76
Db 38 GIKRTPLAKMSVGRKAKASDLDQASVPSBERNSBSSEKRTSDDPFPER--KXA 95
Qy 77 TTVRRG-----KNQVPSLEDSSEPTSTVTVETASBGSVSSSEIRSGPVSDSLCK 130
Db 96 VRAPRGGLGGRKKKAAVASDSKADSDGAKPVPAMARASASSSSSSSDSDVSVK 155
Qy 131 EHPASSB-----KAKGGBEED-----TSDSDGL-TLKEIQNR-----LRR 167
Db 156 KPRGRKPAKRLPFRGRKRPFRPSSSSSDSDSDVDRISBKRDEARRELEARR 215
Qy 168 KRGEPVRSIRGSONRLRKRRREDSABTGSVOIGSA-----BDRPLCKQBPXA- 218

```

```

Db 216 RRGGBEELRLRLEQKEKERRRRR--ADRGAEKSGSGSSGDBLEBDEPVRKGRGR 273
Qy 219 SGPVSGSETDDIENQLEKATQGTENPREAGK--PKPECEYDPMALYICIQPHNN 276
Db 274 GRGPSSSDSBR--BAELR-----REAKSACKPOSSSTEP-----AKRP--- 311
Qy 277 RPMICDRCEWPHDDCVGISRAGRLEBNGEDYICNPCTILQVQDETNGSATNEQDSG 336
Db 312 -----GQEKERYR-----PBEKQ 325
Qy 337 CRVYAGDGTCTSIGTVBOKSGBDGIGRIEKAAPSGKKLK-----IFQVYBAAG 390
Db 396 AKPVYVERT-----RKSGSGSMRKRYEKKGPVBERLOQLHSIRKALKVDSPD 376
Qy 391 APKCIGPCSSVAQPDVYCSNDCTIKR--AAATYRPLSSGEGQTKPKREKVTPEKFS 448
Db 428 TRLSRVLPPIEAVQVYKAKMEKAKBEKLAGEBLAGEBAPOKGAQAGHRSLSPSB 487
Qy 501 W 501
Db 488 W 488

```

RESULT 7 US-09-925-299-862

```

; Sequence 862, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862

```

Query Match 5.9%; Score 190.5; DB 10; Length 518;

Best Local Similarity 22.0%; Pred. No. 2.2e-05;

Matches 119; Conservative 80; Mismatches 185; Indels 157; Gaps 25;

```

Qy 28 GFRRT-----IARBGAGTDEAD-----PSEQPOQHNLRLRSGRQPKTERVBEFL 76
Db 38 GIKRTPLAKMSVGRKAKASDLDQASVPSBERNSBSSEKRTSDDPFPER--KXA 95
Qy 77 TTVRRG-----KNQVPSLEDSSEPTSTVTVETASBGSVSSSEIRSGPVSDSLCK 130
Db 96 VRAPRGGLGGRKKKAAVASDSKADSDGAKPVPAMARASASSSSSSSDSDVSVK 155
Qy 131 EHPASSB-----KAKGGBEED-----TSDSDGL-TLKEIQNR-----LRR 167
Db 156 KPRGRKPAKRLPFRGRKRPFRPSSSSSDSDSDVDRISBKRDEARRELEARR 215
Qy 168 KRGEPVRSIRGSONRLRKRRREDSABTGSVOIGSA-----BDRPLCKQBPXA- 218
Db 216 RRGGBEELRLRLEQKEKERRRRR--ADRGAEKSGSGSSGDBLEBDEPVRKGRGR 273

```

```

: PRIOR FILING DATE: 2001-01-19
: PRIOR APPLICATION NUMBER: 60/263,568
: PRIOR FILING DATE: 2001-01-24
: PRIOR APPLICATION NUMBER: 60/263,694
: PRIOR FILING DATE: 2001-01-24
: NUMBER OF SEQ ID NOS: 155
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 16
: LENGTH: 667
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-863-776-16

Query Match      5.9%; Score 190.5; DB 10; Length 667;
Best Local Similarity 22.5%; Pred. No. 3,1e-05;
Matches 120; Conservative 84; Mismatches 179; Indels 151; Gaps 25;

Dy 28 GFRRTT-----LAKREAGDITDAD-----PSECOPOQHNLISLRSGQPKTERVERFL 76
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 154 GLKRTKTKALKVSYSKAKRAKASDLDQASVPSPEEENSSSEBEKTSDDQPTPEK--KAA 211
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 77 TVVRRG-----KQNVPSLEDSSEPTSTVTVDVTASGVSSESIIRSGVPSDLAG 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 212 VPARRRPPLGGKKGKAPASDSDSKADSGAKPEPAMARSASSSSSSSDSDSVK 271
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 131 EHPASS-----KAKGSEED-----TSQSDSLG--TLKELQNR-----LRR 167
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 272 KPPGRGPAPAKPLPKPGRGPKPERPPSSSSSDSDSDVDRISEWKKRDBARRRLEARR 331
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 168 KRREPEPERSLNGSNNLKKRRBEDSAFTGSVOIGSA-----EODRPLCKQBEPA--SQG 221
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 332 RREDEBELRLRLR--EQEKERKERRRRRAAKNGSG--GSSGELARBDPEPVKKRGKRGGRG 388
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 222 PVSGSEETDLENQLEKATQANTEBNPREAGK--PXPCEVYPNALYICICROPNNRFM 279
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 389 PPSSSDSEP--EALR-----REAKKSAKQSSSTEP-----ARKP----- 423
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 280 ICCDRCEBWHGDCVGISEARGRLERNEDYICPNCTIIQVQDETNGSATNBDSCRS 339
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 424 -----CQEKRRVR-----PEKQQAKP 440
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 340 VGADGTCTSIGTWBKSGSDOGIKGRIEKAAMPGGKKLK-----IFQPVVAPGAPK 393
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 441 VKVART-----RRSSGFSMDKRVKKKPSYBEKQKLHSEIKPLAKYDSPVKR 491
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 394 CIGRCSSVAPQDPVYCSNDCILKH--AAATMFLISGKEOKTKPKKVKYTKDEKPSLPK 451
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 492 CL-----NALELGTLQVTSQILOKNTDVATLK-----KIRRKANDVWEKAAEYVTRL 542
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 452 CSVQVGIKIISVHK--RLASERKNPVKCVMLASRFSGTSGEAPACESSTPSWASD 504
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dy 543 KSRVLAGPIKBAVQKVDKAGMEKEKAERKL---AGBELAGGEAPQEKAKEDKPSSTD 593
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-863-776-18
: Sequence 18, Application US/09863776
: Publication No. US20030198953A1
: GENERAL INFORMATION:
: APPLICANT: Spytek, Kimberly A
: APPLICANT: Majumder, Kumud
: APPLICANT: Tchernyev, Velizar T
: APPLICANT: Mishra, Vishnu
: APPLICANT: Padigaru, Muralidhara
: APPLICANT: Spoderma, Steven K
: APPLICANT: Shenoy, Suresh G
: APPLICANT: Rastelli, Luca
: APPLICANT: Li, Li
: APPLICANT: Taupier, Raymond J
: APPLICANT: Gangoli, Beha
: TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 21402-020
: CURRENT APPLICATION NUMBER: US/09/863, 776

```

CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 09/540,763
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/206,679
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,688
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,829
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,748
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/207,798
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/208,263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/208,831
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 60/209,451
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/210,060
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: 60/219,507
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/221,337
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/221,927
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 60/263,135
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,688
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,694
PRIOR FILING DATE: 2001-01-24
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 667
TYPE: PRT
ORGANISM: Homo sapiens
US-09-863-776-18

Query Match 5.9%; Score 190.5; DB 10; Length 667;
Best Local Similarity 22.5%; Pred. No. 3.1e-05;
Matches 120; Conservative 84; Mismatches 179; Indels 151; Gaps 25;

28 GFRRTT-----IAKREGAGDTAD-----PSEQOPQOHNLSRSGKQPKTERVEEFL 76
154 GLKRTPLALMSVSKARASDLDQASVSPSEENSSSESEKTSDDPTPEK--KAA 211
77 TTVRRRG-----KKNVPVSLDSEPTSTVTDTAEGSVESSESEIRSGPVSDSLK 130
212 VAPRRGPGKGRKKKAPASDSKADSDGAKPEFVNAARASSSSSSSSDSDSVYK 271
131 EHPASSE-----KAKGSEED-----TSDSDGL-TLKELONR-----LRR 167
272 KPRGRKPAEKPLPKPRGRKPKPRPPSSSSSDSDSDDEVDRISSEWRROBARRELEARR 331
168 KROEPVERSLGSGONRLKRRBEDSAETGVOGSA-----EODRPLCKQPEA-SQG 221
332 RROGEELRLRL-BOHEKERRRRRAARSGS--GSSGDELEDEDEBPYKGRKGRGK 388
222 PVSQSETDDIENOLBEKATQGTENPREAGK--PRCEVYDPMALYICROPHNNRM 279
369 PPSSSDSEB-EABLR-----RKAQKSAKPPQSSSTP-----ARKP-----423
280 ICCDRCEWPHGDCVGISBARGLLERNGEDYICPNTLLQVODETNGSATWQDSGCRS 339
424 -----GQKEKVR-----PEKQQAAP 440
340 VADGDTCTSIGTVEKSGEDGIGRIEYANPSGKKLK-----IFQPVYEAAPG 393
441 VVVERT-----RKRSBGFSMDRKVKKKPEVSEKLOKLHSEIKALVDSIDVR 491

394 CIGPCSSVADPDSSVCSNDCLIKH--AAATRFILSGEOKTEKEKVKTERPSLKP 451
492 CL-----NALEIGTLQVTSQILQKNTDVATLK-----KIRRYKANKOVMEGAAYTTL 542
452 CSVVGIIKISSYK-RLASKEKENPVKVMLASRSETSGKEAACSSSTPSMASD 504
543 KSRVIGPTEAVQKVDKAGMEKEKAEKL-----AGBELAGBAPQKABDPSTD 593

RESULT 10
US-10-106-698-5847

Sequence 5847, Application US/10106698
Publication No. US20030109690A1

GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
FILE REFERENCE: PA005PI
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 5847
LENGTH: 555
TYPE: PRT
ORGANISM: Homo sapiens
US-10-106-698-5847

Query Match 5.8%; Score 188.5; DB 14; Length 555;
Best Local Similarity 21.6%; Pred. No. 3.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

28 GFRRTT-----IAKREGAGDTAD-----PSEQOPQOHNLSRSGKQPKTERVEEFL 76
38 GLKRTPLALMSVSKARASDLDQASVSPSEENSSSESEKTSDDPTPEK--KAA 95
77 TTVRRRG-----KKNVPVSLDSEPTSTVTDTAEGSVESSESEIRSGPVSDSLK 130
96 VAPRRGPGKGRKKKAPASDSKADSDGAKPEFVNAARASSSSSSSSDSDSVYK 155
131 EHPASSE-----KAKGSEED-----TSDSDGL-TLKELONR-----LRR 167
156 KPRGRKPAEKPLPKPRGRKPKPRPPSSSSSDSDSDDEVDRISSEWRROBARRELEARR 215
168 KROEPVERSLGSGONRLKRRBEDSAETGVOGSA-----EODRPLCKQPEA- 218
216 RROGEELRLRLBOHEKERRRR--ADRGEARSGSGSDELEDEDEBPYKGRKGRK 273
219 PVSQSETDDIENOLBEKATQGTENPREAGK--PRCEVYDPMALYICROPHNN 276
274 GRGPSSDSEB-EABLR-----RKAQKSAKPPQSSSTP-----ARKP-----311
277 RRMICCDRCCEWPHGDCVGISBARGLLERNGEDYICPNTLLQVODETNGSATWQDSG 336
312 -----GQKEKVR-----PEKQQAAP 325
337 CRVAGDGTCTSIGTVEKSGEDGIGRIEYANPSGKKLK-----IFQPVYEAAPG 390
326 AKPVKVERT-----RKRSBGFSMDRKVKKKPEVSEKLOKLHSEIKALVDSID 376
391 APKCTIGPGSSVADPDSSVCSNDCLIKH--AAATRFILSGEOKTEKEKVKTERPS 448
377 VKKCL-----NALEIGTLQVTSQILQKNTDVATLK-----KIRRYKANKOVMEKAAYV 427
449 LKRCSSVQVGIKISSYK-RLASKEKENPVKVMLASRSETSGKEAACSSSTPSMASD 504
428 TLKRSVLGAPKLEAVQKVDKAGMEKEKAEKL-----AGBELAGBAPQKABDPSTD 481

RESULT 11

US-09-823-187-86
Sequence 86, Application US/09823187
Publication No. US20030096952A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine
APPLICANT: Gusev, Vladimir Y
APPLICANT: Liu, Xiaohong
APPLICANT: Majumder, Kumud
APPLICANT: Padigar, Muralidhar
APPLICANT: Patnirajan, Meera
APPLICANT: Shimeke, Richard A
APPLICANT: Spaderna, Steven K
APPLICANT: Spytek, Kimberly
APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 86
LENGTH: 670
TYPE: PRT
ORGANISM: Homo sapiens
US-09-823-187-86

Query Match 5.8%; Score 188.5; DB 10; Length 670;
Best Local Similarity 21.6%; Pred. No. 4.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

28 GRRRTT-----IARRGAGDTREAD-----PSEQOQOHNLSLRSGRQPKTERREBEL 76
154 GIKRRTKPALMWSVSKARRASDLDQASVSPSEERSSSSSSBKTSDDPTERK--KAA 211
77 TTVRRG-----KKNVPSLEDSSEPTSTVDTVTAISGVSSEISRGVPSISLG 130
212 VRAPRGPIAGRGKKAAPSASDSSTADSDGAKPEVPAARASSSSSSSSSSDSDSVSK 271
131 EHPASS-----KAKGHEBED-----TSDSDSDGL-TLKEIQNR-----LRR 167
272 KPRGRKAPLEKPLPKRGRKPKRPRSSSSSSSDSDSDVDRISEWKRDRARBLEARR 331
168 KBOEPVERSLKSGONRLAKGRREBSAETGVSQIGSA-----EQRPPLKQDEPA 218
332 RREQEBELRLRLEQBEKBEKRRER--ADRGAEKSSGSSGDELRRDDPEPVKRGKGR 389
219 SGGVPSQSHTDIENLEGATQANTENRBAK--PRPECEVDPNALYCICROPHN 276
390 GNGPSSSSSEPP-EAELR-----RAKKAKAKPQSSTEP-----AKRP--- 427
277 RMFICCDRCBWFHGDVGISBARGLLRNEDYICPNCTIIQVDETNGSATNEQDSG 336

RESULT 12

US-09-863-776-51
Sequence 51, Application US/09863776
Publication No. US20030198953A1
GENERAL INFORMATION:
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Padigar, Muralidhara
APPLICANT: Spaderna, Steven K
APPLICANT: Shenooy, Suresh G
APPLICANT: Rastelli, Luca
APPLICANT: Li, Li
APPLICANT: Taupier, Raymond J
APPLICANT: Gangcilli, Beba
TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-020
CURRENT APPLICATION NUMBER: US/09/863,776
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 09/540,763
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/206,679
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,688
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,829
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,748
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/207,798
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/208,263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/208,831
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 60/209,451
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/210,060
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: 60/219,507
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/221,337
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/221,927
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 60/263,135
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,688
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,694
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 51
LENGTH: 670
TYPE: PRT

428 -----GQEKRYR-----PEEKQ 441
337 CRVAGADGTCTSIGTVBQSGEDQGIKRIKAAKNSKKKXK-----IFQPVYAPG 330
442 ACPVVERI-----RKRGSEFSMDRKVEKKEKPSVEKGLQKHSEIKFKALTKDSDPD 492
391 APKCIGPGCSSVAGPDSVYCSNDCLIKH--AAATRFPLSSGKQKTKPKKXVTKREKPS 448
493 VRCL-----NALREIGTLQVTSQILQKNTDVVAATK-----KIRRYKANDVNEKAAEVY 543
449 LKPCSVQVGIKISVYK-RLASEKRENPKVWMLASRSETSKGAAACSSSTPSMSAD 504
544 TLKSKRVLPKTIKAVQKVKAKGHEKKEKAEKL---AGEELAGEAPOKRAEDKDPSTD 597

ORGANISM: Homo sapiens
US-09-863-776-51

Query Match 5.8%; Score 188.5; DB 10; Length 670;
Best Local Similarity 21.6%; Pred. No. 4.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

```
QY 28 GRRRTT-----IAKREGAGDTAD-----PSEQPOQHNLILKRGKPKTERVERBL 76
DB 154 GIKRTIPALMKNVSKARASDLQASVSPSESESSSSSEKTSDDFTPEK--KAA 211
QY 77 TVVRRG-----KKNVPVSLDSSSEPTSTVTDVETASRGVSSSEBIRSGVPVSLCK 130
DB 212 VAPRRGLGRKKKAPASDSKADSDGAKPEVVAARSSSSSSSSSSSSSDSVSK 271
QY 131 EHPASSR-----XAKGEBRED-----TSDSDSGL-TLKEQLNR-----LRR 167
DB 272 KPRGRKPAKPLPKPRGRKPKPRPPSSSSSDSDSDVDRISEWKRREARRELEARR 331
QY 168 KKEQEPVERSLGSONRLKKRREDSATGVSVOGSA-----EODRPLCKQPEA- 218
DB 332 RREQEBELRLRBOEKERERRER--ADRGELRSGSGSGDELNEDPEVKRGRKGR 389
QY 219 SGGPSSQSDTDIENLBKATQGTNEEMPREAK--PKPRCEVVPNALYICRQPHNN 276
DB 390 GGGPSSSDSEP-EKELR-----REKKSAKPKQSSSTP-----AKP----- 427
QY 277 RFMICCDRCCEWPHGDCVGISBARGRLERNGBDYICPCTILQVQDETNGSATNEDQSG 336
DB 428 -----GQKEKRVX-----PREKQ 441
QY 337 CRVSGAGDICTSIGTVEOKSGHGOIKRIEKAAPSGKKLTK-----IPQVYEAAG 390
DB 442 ARPVKVET-----RKRSQFSGMDRKVEKKKPPVEEKLQKXLSBKFLKXVDSPD 492
QY 392 APKCIQPGSSVAQPSVYCSNDCLKH--AAATNPLFSGKQKTKPKKVKTKDEKS 448
DB 493 VKRCL---NALBELGTLOVTSQILOKXNDVYATLK-----KIRKANKDVWEKAAEY 543
QY 449 LPKCSVQVGIKISVYHK-RLASERENPVKCVMLASRSTSGKAAECSSSTPSWASD 504
DB 544 TLKSRVGLGPKLEAVGVKNKAKMEKEKAEKL--AGBELAGBEAQEKAKEDKPSFD 597
```

RESULT 13

US-09-946-374-308

Sequence 308, Application US/09946374

Publication No. US20030073129A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc
APPLICANT: Batton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Auestin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paonl, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumes, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
ACIDS
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C1

CURRENT APPLICATION NUMBER: US/09/946,374
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930

PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103396
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103401
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103449
PRIOR FILING DATE: 1998-10-06

PRIOR APPLICATION NUMBER: 60/103633
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807

Query Match 5.8%; Score 188.5; DB 10; Length 671;
Best Local Similarity 21.6%; Pred. No. 4,3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

QY 28 GPRRTT-----IAKRGAGDTAD-----PSEQPOQHNLSLRSGROPKTERVEEFL 76
DB 154 GLKRTPLAKMSVSRARAKASDLQASVSPSEESSESSSEKTSDDFTPEK--KAA 211
QY 77 TTVRRRG-----KQNPVLSLEDSSEPTSTYTDVETASGCVSESEIRSGPVSDSLK 130
DB 212 VAPRRGPGKGRKKKAPASDSDSKADSDGAKPEVAMARASASSSSSSSDSVSVK 211
QY 131 EHPASSE-----YAKGGBREED-----TSDSDSDGL-TLKEIQNR-----LNR 167
DB 272 KPPRGKRPKRPKPLPRKPKPKPRPPSSSSSDSDSDVDRISWKRRRDEARRRELEARR 331
QY 168 KREQBPVERSLRSGQNRLEKKREEDSAETGSAVQIGSA-----EQRPLCKQPEA- 218
DB 332 RRDQEEELRLRQEKREKRRRR--ADRGALRSGSGSSGDELREDDPEVKRGRKGR 389
QY 219 SGGPVQSSTDIENOLBKATQANTENPREAGK--PPDECVVDPNMLYCIQROPHN 276
DB 390 GRGPPSSDSRP-EARLE-----REAKSAKTPQSSSTP-----ARKP--- 427
QY 277 RFNCCDRCEBWFHGDGVGISBARGLLEKNGEDYICPNCTLLQVODETNGSATNEQDSC 336
DB 428 -----GQKRYV-----PEKQ 441
QY 337 CSVAGDGTCTSIGTVQKSGEDGIGRIKXANPSEKTKK-----TFQPVYAPG 390
DB 442 AKPVKVER-----RKRSQFSMDRKVEKKEPSVEKQKXLSHSEIKFALKYDSPD 492
QY 391 APKCIQPGCSSVAQPSVYCSNDCLIKH--AAATAPLSSGKQKPKPKXVTKPEKFS 448
DB 493 VRKCL-----NALRELTQVTSQILQKNTDVATLK-----KIRRYKANDVKEKALEVY 543
QY 449 LPKCSVQVGIKISVYK--RLASBKRENPKVQWLASRSETSGKAAECSTPSPASD 504
DB 544 TRLSKRVLPKILAVQKVNKAKGKREKARKKL--AGEELAGBRAQKAEADPSPD 597

RESULT 14
US-09-823-187-85
; Sequence 85, Application US/09823187
; Publication No. US20030096952A1

GENERAL INFORMATION:
APPLICANT: Burgess, Catherine
APPLICANT: Gusev, Vladimir Y
APPLICANT: Liu, Xiaohong
APPLICANT: Majumder, Kumud
APPLICANT: Padigar, Muralidhar
APPLICANT: Patturajan, Meera
APPLICANT: Shinkens, Richard A
APPLICANT: Spaderna, Steven K
APPLICANT: Spytek, Kimberly
APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
PRIOR FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 85
LENGTH: 671
TYPE: PRT
ORGANISM: Homo sapiens
US-09-823-187-85

Query Match 5.8%; Score 188.5; DB 10; Length 671;
Best Local Similarity 21.6%; Pred. No. 4.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

28 GPRATT-----IAKREGAGDTEAD-----PSEOPFOCHNLILRRSGRQPKTERVEREPL 76
154 GUKRTTPALKMSVSKARAKASDLDQASVSEERNSSESSSEKTSDDPTPER--KAA 211
77 TTVRRRG-----KKNVPVSLDSSEPTSTVTVTETASSEGVSSSEIRSGPVSDSLAK 130
212 VRAPRGPGKGRKKKAPASDSDSKADSDAKPEPVAMARASSSSSSSSDSDSVYK 271
111 EHPASSF-----KAKGSEED-----TSDSDSLG-TLKELONR-----LBR 167
272 KPRGRKPAEKPLPFRGKPKPRPPSSSSSDSDSDVDRISEWKRREARRELEAKR 331
168 KKEQEPVRSLLSGONRLAKKRREDSAVTGSVQIGSA-----EODRPLCKQEPFA- 218
332 RREGEELRLRLBQKEKERERRR--ADRGABRNGSGSSSDDELREDEPVRKGRKR 389
219 SGGPVQSSETDDIENQLEGKATQNTENPREAGK--PRGCEVYDPAALYICICROPHNN 276
390 GRGPSSSSSEF--EAELE-----REAKSKAKPKOSSSTEP-----ARKP--- 427
277 RRMICDRCBEMFHDDCVGISARGLLEARNEDYICPCTTLQVQDDETNGSATNEQDSG 336
428 -----GQKSKRVR-----PEKKQ 441
337 CSRVGADTDCSTIGTVEQSGEDGQIKRIEKAANPSGKKLX-----IFQPVVEAPG 390

442 AKPVKERT-----RKSESGPSMDRKVEKKKPSVEEKLQKLHSEIKELKALVDSBD 492
QY 391 APKCIIGPCSSVAOPDSVYCSNDCLKH--AAATRFPLSSGEOKPKPEKTKPEKPS 448
DB 493 VKRCL-----NALBEIGTLQVTSQILQKNTDVATLK-----KIRYKANKDVMEKAAEY 543
QY 449 LPKCSVQYIKISVHK--RLASERENPVKVMLASRSETSGKEAACESSTPSMSD 504
DB 544 TRKSRVIGPFIENQYKNKQKMEKKAEBKL---AGSELAGEBAPOEVAIDKPSFD 597

RESULT 15
US-09-863-776-20
Sequence 20, Application US/09863776
Publication No. US20030198953A1
GENERAL INFORMATION:
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Padigar, Muralidhar
APPLICANT: Spaderna, Steven K
APPLICANT: Shenoy, Suresh G
APPLICANT: Rastelli, Luca
APPLICANT: Li, Li
APPLICANT: Taupier, Raymond J
APPLICANT: Gargolli, Esba
TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-020
CURRENT APPLICATION NUMBER: US/09/863,776
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 09/540,763
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/206,679
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,688
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,829
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,748
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/207,798
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/208,263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/208,831
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 60/209,451
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/210,060
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: 60/219,507
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/221,337
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/221,927
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 60/263,135
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,688
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,694
PRIOR FILING DATE: 2001-01-24
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 671
TYPE: PRT
ORGANISM: Homo sapiens
US-09-863-776-20

Query Match 5.8%; Score 188.5; DB 10; Length 671;
Best Local Similarity 21.6%; Pred. No. 4.3e-05;

Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

QY 28 GFRRTT-----IAKRGAGCTPDA-----PSEQPQOHNLIRSGOPKTEREETL 76
 Db 154 GLKRTKPTALNWSVSKRARKASDLDQASVSPSEENSSSSSEKSTODDFTPEK--KAA 211
 QY 77 TTVRRRG-----KQNVPSLEDSSEPTSTVTVDBTASGEGVSSSESTRSGPVSILK 130
 Db 212 VRAPRKPLGGRKKKKAPASDSDSKADSDGAKPEVMAKARASSSSSSSSSDSVSVK 271
 QY 131 EHPASSR-----KAKGGEED-----TSDSDGL-TLKELONR-----LIR 167
 Db 272 KPRGGRPAKEKELPKRGKPKPKPPSPSSSSSDSDSDVEDRISWWRBROEARRLEARR 331
 QY 168 KRQEPVPSRLSGSONRLKTKRREDSALTSVQIGSA-----EDRPLCKQEPRA- 218
 Db 332 RRQEBEIRLRLBQBEKERRRRR--ADRGAEKSGSGSDGLREDEPVKRRKRR 389
 QY 219 SGQPVSGSETDDINQLEGKATQGTENPPRAGK--PRQGEVDPNMLYCIQROPHNN 276
 Db 390 GRQPPSSSDSEP-BAELE-----RKAKSAAKPPSSSTEP-----ARKP- 427
 QY 277 RFMICCRCEBMFHGDVGISEARGRLERNGEDYICPCTILLQVQDETNGSATNEQDSG 336
 Db 428 -----GQKEKVR-----PEKKQ 441
 QY 337 CRVSGADGDTCTSIGTVEQKSGEDQIGRIERKANPSGKKLK-----TFQPVYAFG 390
 Db 442 AKPKVKERT-----RKSGEFSMDRKVEKKKEPVEEKLQKLSSEIKFALKVDSPD 492
 QY 391 APKCTIGECSSVAOPDSVYCSNDCTLKH--AAATYRFLSSGSGQKTKPKRYKTRBKFS 448
 Db 493 VKRCL-----NALBEIGTLQVTSQILQKNTDVAAVLK-----KIRRYTAKNDWEKAAEYV 543
 QY 449 LPKSVQVGIYIISVHK-RLASKEKENPVKMYLMSRSTSGKBAACSSSTPSMAAD 504
 Db 544 TRLSRVLGPPIBAVQVKNKAMEKKEEKU--AGEBLAGEBAOEVAEDKPESTD 597

Search completed: April 20, 2004, 21:25:48
Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:17:09 ; Search time 23 Seconds
(without alignments)
1378.189 Million cell updates/sec

Title: US-09-787-016a-4
Perfect score: 3228
Sequence: 1 MDDGHLNBERAPKIKRPTS.....LKSCVGLMLNLSYRFRPW 614

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptcodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptcodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptcodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptcodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptcodata/2/1aa/6CTUS_COMB.pep:*
6: /cgn2_6/ptcodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	197.5	6.1	1805	1	US-07-853-913-2 Sequence 2, Appl
2	166	5.1	688	3	US-09-141-047-8 Sequence 8, Appl
3	165.5	5.1	1780	1	US-08-769-309A-5 Sequence 5, Appl
4	165.5	5.1	1780	3	US-08-994-570-5 Sequence 5, Appl
5	164	5.1	238	4	US-09-257-179-80 Sequence 80, Appl
6	153.5	4.8	1848	4	US-08-296-791-6 Sequence 6, Appl
7	153.5	4.8	1848	4	US-09-839-996-6 Sequence 6, Appl
8	153.5	4.8	1848	4	US-10-080-505-6 Sequence 6, Appl
9	153.5	4.8	1848	5	PCT-US95-10651A-6 Sequence 6, Appl
10	151.5	4.7	532	1	US-08-285-440-5 Sequence 5, Appl
11	151.5	4.7	532	1	US-08-630-349-5 Sequence 5, Appl
12	149.5	4.6	2468	4	US-09-976-594-726 Sequence 726, App
13	148	4.6	657	3	US-08-893-852A-3 Sequence 3, Appl
14	148	4.6	657	3	US-08-821-818-3 Sequence 3, Appl
15	148	4.6	657	4	US-09-052-753B-3 Sequence 3, Appl
16	144.5	4.5	1088	4	US-09-130-242-2 Sequence 2, Appl
17	142	4.4	1596	4	US-08-978-277A-4 Sequence 4, Appl
18	141.5	4.4	1739	4	US-09-976-594-76 Sequence 76, Appl
19	141	4.4	8991	4	US-08-714-741-32 Sequence 32, Appl
20	140.5	4.4	558	1	US-08-285-440-6 Sequence 6, Appl
21	140.5	4.4	558	1	US-08-630-349-6 Sequence 6, Appl
22	138.5	4.3	2289	3	US-09-051-019-2 Sequence 2, Appl
23	135.5	4.2	1233	4	US-09-645-456A-35 Sequence 35, Appl
24	135.5	4.2	1233	4	US-09-425-324A-35 Sequence 35, Appl
25	135.5	4.2	1233	4	US-09-645-791-35 Sequence 35, Appl
26	134.5	4.2	1231	4	US-09-595-684B-23 Sequence 23, Appl
27	134.5	4.2	1297	4	US-09-688-188B-14 Sequence 14, Appl

28	134.5	4.2	1297	4	US-09-291-417D-14 Sequence 14, Appl
29	134.5	4.2	1332	4	US-09-645-456A-9 Sequence 9, Appl
30	134.5	4.2	1332	4	US-09-425-324A-9 Sequence 9, Appl
31	134.5	4.2	1332	4	US-09-645-791-9 Sequence 9, Appl
32	134	4.2	1075	4	US-09-252-991A-18387 Sequence 18387, A
33	134	4.2	3256	4	US-09-919-172-98 Sequence 98, Appl
34	134	4.2	3256	4	US-09-976-594-22 Sequence 22, Appl
35	132.5	4.1	491	1	US-07-903-103-2 Sequence 2, Appl
36	132.5	4.1	491	1	US-08-044-619A-2 Sequence 2, Appl
37	132.5	4.1	491	1	US-08-283-911-2 Sequence 2, Appl
38	132.5	4.1	491	1	US-08-245-500A-3 Sequence 3, Appl
39	132.5	4.1	491	1	US-08-390-546-3 Sequence 3, Appl
40	132.5	4.1	491	1	US-08-390-479A-3 Sequence 3, Appl
41	132.5	4.1	491	1	US-08-557-393-3 Sequence 3, Appl
42	132.5	4.1	491	1	US-08-390-516C-3 Sequence 3, Appl
43	132.5	4.1	491	1	US-08-390-517A-3 Sequence 3, Appl
44	132.5	4.1	491	1	US-08-390-515A-3 Sequence 3, Appl
45	132.5	4.1	491	2	US-08-801-718-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-07-853-913-2
Sequence 2, Application US/07853913
Patent No. 5338639
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.
TITLE OF INVENTION: Nestin Expression As An Indicator of
TUMOR OR INFECTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein

US-07-853-913-2

Query Match 6.1%; Score 197.5; DB 1; Length 1805;
 Best Local Similarity 21.2%; Pred. No. 8.3e-08;
 Matches 143; Conservative 99; Mismatches 223; Indels 211; Gaps 35;

QY 2 DDKG--HLSNEBAPKAIPTSKKFRRTTIAKRGAGDTADPSQQPOQHLS 58
 DB 486 EDEGQIWEIWEKADIEVVENSSAQT-----QESGLDTETDSDQPLQKR-T 534
 QY 59 LARSROPRTREVEREFLTVRRGKNVPVSLSDS-----SEPTSTVTIDVERASGSYE 114
 DB 535 LVALBEPPLMSLTKQVETRA-----GKENCNSTBGLGLTBGPBKQIPLSLKKNV 590
 QY 115 SSSSEIRSG-PV-SDSISKEHPASSER-----AKG-----GEEEDTSDSDG----- 155
 DB 591 SKRTLENGVPVLSLGLKEDRTEDDELMSPGKTLRFPSLSGSGEVAVRPSKEGMLBSM 650
 QY 156 -----LTLKELQN-----RLRRKRGQR----- 173
 DB 651 TAFKESQHPGLGPGABDQMLRLVKEQDSPPRSPREBDOBAQPLQKENDPEPLGYEBA 710
 QY 174 -----VERSL-RGSQNRRLRKGREDSAGTSVQIGSABDRPLCKQEPSPASQSPVSCSET 228
 DB 711 EQQILRLLEKESQESLRSP-----EEEDQENG-----RSLQK-----NQEPLGYEBA 754
 QY 229 DD--IENOLEGKATOG--NTEENPREAGK-----KPECEVYDPMALYCTICQPHNNRPM 279
 DB 755 EDQMLERLLEKESQESLKGPEEMQR--IGKPLERENQSLRYLEENQBTVPPLSRNQPL 813
 QY 280 ICCDRCER-----WPHQCVGIS-----EARGLU 303
 DB 814 RSLVEEBEQRIVKPLKVSQDSLSGLAEENQPLRYLEBDICINSLBDTKTKSLGSL 873
 QY 304 LBRNGEDYICPNCTIIQV-----ODETNGSATNEODSGCRSVGADGTD 346
 DB 874 EDRNGDSIIIPQBSERQVSLRPPBEDQRIYNHLEKESQEPSPRSSEBQWMRSLSEGN 933
 QY 347 CTSIGTVEKSGEDQIKGRILEKANPSPGK-----KLLKIPQVVEAPGAPKICIGPCSSV 402
 DB 934 HESLSLSEVER--EDQVSESQLEKESQDSGKSLDESDQETGRL-----EKNALESLSL 984
 QY 403 AGPDVYCCNDCLIKLAATMRPLSSGKQ-KTKPEKXKTKPEKSLPKCSVQVQIKIS 461
 DB 985 AGQDQ--EBQKLEQETQOTLR--AVGNEQWAVSPPEKV--DPR--PLG--NDQ 1030
 QY 462 SVHKRLASEKEN-----PVKKV-----MLASRSETSGKAACGSPSPVNASDH 505
 DB 1031 ELARSLGKNGESVLVLEKGIETVKSLETRIIETLETAEDLERKKSIDTOEPPLMSTRV 1090
 QY 506 NYNAVPEKPEKPTAL 521
 DB 1091 ARETVPEPEDEPGSL 1106

RESULT 2
 US-09-141-047-8
 ; Sequence 8, Application US/09141047A
 ; Patent No. 6043085
 ; GENERAL INFORMATION:
 ; APPLICANT: Walker, David H.
 ; APPLICANT: Yu, Xue-Die
 ; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
 ; TITLE OF INVENTION: Protein Gene
 ; FILE REFERENCE: D6143
 ; CURRENT APPLICATION NUMBER: US/09/141,047A
 ; CURRENT FILING DATE: 1998-08-27
 ; NUMBER OF SEQ ID NOS: 15
 ; SEQ ID NO 8
 ; LENGTH: 688
 ; TYPE: PRT
 ; ORGANISM: Ehrlichia canis
 ; FEATURE:

OTHER INFORMATION: Amino acid sequence of 120 kDa immunoreactive
 ; OTHER INFORMATION: protein.
 ; US-09-141-047-8

Query Match 5.1%; Score 166; DB 3; Length 688;
 Best Local Similarity 21.4%; Pred. No. 1.1e-05;
 Matches 128; Conservative 84; Mismatches 211; Indels 176; Gaps 30;

QY 8 SNEBAPKAIPTSKKFRRTTIAKRGAGDTADPSQQPOQHLSLRSGROPK 67
 DB 90 SSSSEVKKVSKTSK-----ESTPEVKA-----DLQPAVDGSIHSSS--EVGKVS 135
 QY 68 KTERVEEFLTVRRGKNVPVSLSDSSEPTSTV-----TVETAS 109
 DB 136 KTSK--ESTPEVK--ABDLQPAVDGSIHSSSEVGEKVSSTSKENTPEVKAEDLOPAV 191
 QY 110 EGSVE--SSSEI-----RSGP-----VSDSLGKHPAS--SEKAKGEBER 146
 DB 192 DSGIHSSEVGEKVSSTSKENTPEVKAEDLOPAVDGSI--EHSSEVGEKVSSTSKER 249
 QY 147 DTSDSDGCLTLKELQNLRLRRGQEPVERSLAGSQNLKTKRREDSAPTSVQIGSAR 206
 DB 250 NTEPVVABDL-----QPAVDGSIHSSSEVGEK--VSETSKESTPEV-KAR 293
 QY 207 QDRPLCKQBPASQSPVSO--SETDIEKOLEGK-----ATQANTEENPREAG----- 252
 DB 294 DLQPAVDGSIHSSSEVGEKVSSTSKENTPEVKAEDLOPAVDGSIHSSSEVGEKVSST 353
 QY 253 ---KPECEVYDPMALYCTICQPHNNRPMICCDRCBEMFHQDQVGISEARGLLERNGB 309
 DB 354 SKESTPEVKAEDLOPAVDGSIHSSS--LQ-----AVDSIR--HS--SSEVGEKVSSTSK 391
 QY 310 DYICPNCTIIQVDETFNGSATNEODSGCRSVGADGTDCTISIGVBOKSGD--QGIKRI 367
 DB 392 ESTPEVKAEDLOPAVDGSIHSSS--EVGKVSSTSKENTPEVKAEDLOPAVDGSI 447
 QY 368 EKANPSPGK-----KLLKIPQVVEAPGAPKICIGPCSSVAPDPSVCSND 414
 DB 448 EHSSEVGEKVSSTSKENTPEVKAEDLOPAV-----DG 481
 QY 415 ILKHAATM--RFLSSGKQKTKPEKVKTKPE--KESLPKCSVQVQIKISSVK-----R 466
 DB 482 SVHSSSEVGEKVSSTSKENTPEVKAEDLOPAVDGSIHSSSEVGEKVSSTSKESTPE 541
 QY 467 LASEKREPMYKVMALASRSTSGK--EACCSSTPSMASDHNTNAYPEKPEKPTALSP 523
 DB 542 VKAEDLOPAVDGSIHSSSEVGEKVSSTSKESTPEVKA-----VQPAVDGMPVPLNP 595

RESULT 3
 US-08-769-309A-5
 ; Sequence 5, Application US/08769309A
 ; Patent No. 5741890
 ; GENERAL INFORMATION:
 ; APPLICANT: Scott, John D.,
 ; APPLICANT: Nauert, Brian J.,
 ; APPLICANT: Klauk, Theresa M.
 ; TITLE OF INVENTION: Protein Binding Domains of Gravin
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gershtein, Murray & Borun
 ; STREET: 6300 Sears Tower/233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/769,309A

FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 574189and, Greta E.
 REGISTRATION NUMBER: 35,302
 REFERENCE/DOCKET NUMBER: 27866/33451
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448
 TELE: 25-3856
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1780 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULAR TYPE: protein
 US-08-769-309A-5

Query Match 5.1%; Score 165.5; DB 1; Length 1780;
 Best Local Similarity 20.4%; Pred. No. 4.8e-05;
 Matches 133; Conservative 81; Mismatches 201; Indels 237; Gaps 32;

QY 3 DKGHLSN--BEAPKA-----IKPTSK-----FRKTM--GPRRT----- 32
 DB 136 DDGQENRNIBOIPSSSMLBELTQPTESQANDIGKFKVGFPTVKQKTERPDTV 195
 QY 33 ---TIKREGAGDTA---DPS-----EQOQOHNLSLRSGRQPKTERVEEFL 76
 DB 196 QLLTVKDEBGAAGAGHDPSLGAAGAAKSEBKKO-----STKKEBETL 242
 QY 77 TVRRRGKKNVPSLDSSEPTSTVTDVETASGSSVSSSIRSGPVSLSLCK----- 130
 DB 243 KRGSHAEISPPA---ESGQAVBCEKGEKPKSKASASPTSPVTSSTGTFKPF 299
 QY 131 -----EHPASSBKAKGGEEDTSDSDGLTLKELNRLRRKROEPEVERSLGSON 183
 DB 300 TQWAGMRKKTSPFRPK--EDVEVASE-----KKKEQBP----- 331
 QY 184 RLKRRKREED-SAETSGVOIGSABDRPLCKQEP--BASQGPVSQSETDI-----ENQLBG 237
 DB 332 --EKVDTEBDGALVASEKLTASBQHP---QEPASAHPRLSATYKVELPSEBQVSG 386
 QY 238 KATQNTBENPBEAGPKP--ECBVYDPMALYCIQRPNNRPMICCDRCBEMFHGDCVGI 296
 DB 387 --SQGPSEB---KPAFLATEVFDEKI-----EVHQBKVVA-----EVHV 420
 QY 297 SEARGHLERNGEDYICNPCTILQVODETNGSANTNQDSCGCSVGADGTDCTSIGTVEOK 356
 DB 421 STVEBTEBQKTE-----VEETAGSVPAEELVGMDAEPQEAEPAKELVTLKET 468
 QY 357 --SGED--OGI-----KGRIRKANPSSGKKLKI-----FQPVVEAPGAPKCI 395
 DB 469 CVSGEDPTQADLSPEDEKVLKRPBGVSVSEVEMLSQGEKMKVGSPLKGLFTSTGLKLS 528
 QY 396 G-----PGSSVAOPDSVYCSNDCILHAAA 421
 DB 529 GKQKQKCGGEBEGENTQVPADSPDSQBPQKSSASSPPEPEITC-----LEKGLA 583
 QY 422 TWRFLLSGKEQTKPKPEKVK---TKPEKPSLPKCSVQVQVGIKISSVHKRLASEGRNP 475
 DB 584 EVQOQGEABEGATSDQEKKBGVTPWASFKKWTTPKGRV-----RRPSBSDKEDB 633
 QY 476 VKKVLASSETSGKEACBSSTPWSADHNANVAPREKPTALSPITLLS 527
 DB 634 LDKVKSATLSST-----BSTASEKQEMKGSVEEPPEPKKRVDTSVS 677

RESULT 4
 US-08-994-570-5
 Sequence 5, Application US/08994570
 Patent No. 6090929
 GENERAL INFORMATION:
 APPLICANT: Scott, John D.,

APPLICANT: Nauert, Brian J.,
 APPLICANT: Klauck, Theresa M.
 TITLE OF INVENTION: Protein Binding Domains of Gravin
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
 STREET: 6300 Sears Tower/233 South Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/994,570

FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6090929and, Greta E.
 REGISTRATION NUMBER: 35,302
 REFERENCE/DOCKET NUMBER: 27866/33451
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448
 TEL: 25-3856

INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1780 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULAR TYPE: protein
 US-08-994-570-5

Query Match 5.1%; Score 165.5; DB 3; Length 1780;
 Best Local Similarity 20.4%; Pred. No. 4.8e-05;
 Matches 133; Conservative 81; Mismatches 201; Indels 237; Gaps 32;

QY 3 DKGHLSN--BEAPKA-----IKPTSK-----FRKTM--GPRRT----- 32
 DB 136 DDGQENRNIBOIPSSSMLBELTQPTESQANDIGKFKVGFPTVKQKTERPDTV 195
 QY 33 ---TIKREGAGDTA---DPS-----EQOQOHNLSLRSGRQPKTERVEEFL 76
 DB 196 QLLTVKDEBGAAGAGHDPSLGAAGAAKSEBKKO-----STKKEBETL 242
 QY 77 TVRRRGKKNVPSLDSSEPTSTVTDVETASGSSVSSSIRSGPVSLSLCK----- 130
 DB 243 KRGSHAEISPPA---ESGQAVBCEKGEKPKSKASASPTSPVTSSTGTFKPF 299
 QY 131 -----EHPASSBKAKGGEEDTSDSDGLTLKELNRLRRKROEPEVERSLGSON 183
 DB 300 TQWAGMRKKTSPFRPK--EDVEVASE-----KKKEQBP----- 331
 QY 184 RLKRRKREED-SAETSGVOIGSABDRPLCKQEP--BASQGPVSQSETDI-----ENQLBG 237
 DB 332 --EKVDTEBDGALVASEKLTASBQHP---QEPASAHPRLSATYKVELPSEBQVSG 386
 QY 238 KATQNTBENPBEAGPKP--ECBVYDPMALYCIQRPNNRPMICCDRCBEMFHGDCVGI 296
 DB 387 --SQGPSEB---KPAFLATEVFDEKI-----EVHQBKVVA-----EVHV 420
 QY 297 SEARGHLERNGEDYICNPCTILQVODETNGSANTNQDSCGCSVGADGTDCTSIGTVEOK 356
 DB 421 STVEBTEBQKTE-----VEETAGSVPAEELVGMDAEPQEAEPAKELVTLKET 468
 QY 357 --SGED--OGI-----KGRIRKANPSSGKKLKI-----FQPVVEAPGAPKCI 395
 DB 469 CVSGEDPTQADLSPEDEKVLKRPBGVSVSEVEMLSQGEKMKVGSPLKGLFTSTGLKLS 528
 QY 396 G-----PGSSVAOPDSVYCSNDCILHAAA 421

Db 529 GKKOKKRGCGDBRSGEHTQVPAIDPSOBKQKSSASSPPEPBITC-----LEKGLA 583
Qy 422 TWRPLSSGKOKTKPEKVK-----TKPEKPSLPKCSYQVVGIKISSVHKRLASEKREP 475
Db 584 EVQOQOABEABGATSDCKRGVTPWASPQWVTPKRV-----RRPEESDKEDB 633
Qy 476 VCKQMLASSBTSKGKACBSSPSPMAASHNRYAVPEPEKPTALSPILLS 527
Db 634 LDKVKSATLST-----ESTASSEMOKSGSVPEPEKPKRVKVTSVS 677

RESULT 5

US-09-257-179-80
Sequence 80, Application US/09257179
Patent No. 6410709
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 29 Human Secreted Proteins
FILE REFERENCE: P2015P1
CURRENT APPLICATION NUMBER: US/09/257,179
EARLIER FILING DATE: 1999-02-25
EARLIER APPLICATION NUMBER: PCT/US98/17709
EARLIER FILING DATE: 1998-08-27
EARLIER APPLICATION NUMBER: 60/056,270
EARLIER FILING DATE: 1997-08-29
EARLIER APPLICATION NUMBER: 60/056,271
EARLIER FILING DATE: 1997-08-29
EARLIER APPLICATION NUMBER: 60/056,247
EARLIER FILING DATE: 1997-08-29
EARLIER APPLICATION NUMBER: 60/056,073
EARLIER FILING DATE: 1997-08-29
NUMBER OF SEQ ID NOS: 128
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 80
LENGTH: 238
TYPE: PRT
ORGANISM: Homo sapiens
US-09-257-179-80

Query Match 5.1%; Score 164; DB 4; Length 238;
Best Local Similarity 36.2%; Pred. No. 3.4e-06;
Matches 29; Conservative 16; Mismatches 11; Indels 24; Gaps 3;

Qy 238 KATQGTENPREAGKPECEVYDPAALYCTROPHN-NRFMICDRCGEWTHGCVGI 296
Db 51 RAQGGSSER-----LYCICRTYDSDQFYIGCDRCQNTWTHGRCVGI 91
Qy 297 SEARGRLERKEDYICPNC 316
Db 92 LOSEABELT-----DEVVCPQC 107

RESULT 6

US-08-296-791-6
Sequence 6, Application US/08296791
Patent No. 624337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSES: Plehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard P.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-6

Query Match 4.8%; Score 153.5; DB 3; Length 1848;
Best Local Similarity 21.7%; Pred. No. 0.00055;
Matches 124; Conservative 71; Mismatches 208; Indels 169; Gaps 25;

Qy 8 SNEEAKAIKPTSKERKTMGFRRTIARREGAGTDEADPSE-----OQPOCHNTSLRR 61
Db 1072 STETAPKSDTATVTE-----NPNSESVPSTETKVAENPPENETVAK 1114
Qy 62 SGRO-----PKTERVEEPLTV-----RRGKNVVPVLSDSHPTSTVTVDR 106
Db 1115 NQGEATEPTPQNEVANKEDQPTVEAVTQNEATQSGKQEFYQAFKSPETS-VTVSG 1173
Qy 107 TASEGVSSESSERTSPVSDSLCKEHPASSEKAKGGEKHDTSDSDGLTKELQRLR 166
Db 1174 MPEKTVSOSTE-----DKVVE-----KREKAK--VETEFYQAKQ--VTSKRPKQAR 1219
Qy 167 RREKQSVERSLGSQNRRLKRRBEDSAFTQSGAQQ-DRLCKQSEPR-----217
Db 1220 PAPERVPTDINAEQAALQOTQPTTVAALFTISPNKSPARETQPSBKTAIEVTVPVSE 1279
Qy 218 --ASQGPVSGSETDIDENQLEGATQNTENPRE--AGKP-----KPECEVYDPAALYC 268
Db 1280 NTATQ-PTETEEYAKYKXKTKQEVQVVASQSPKQDPAKQAYKQAPARENVL-- 1336
Qy 269 ICRQPHNNRPMICDRCGEWTHGDCVIGSEARGLLERNGEDYICNCTILQYQDETNGS 328
Db 1337 -----TKNVGEPO-----POAQPOQST 1355
Qy 329 A---TNEODSGCKSVGADGTC-----TSIGTVBQSGEDQGIKGRIBKANPSGCK 377
Db 1356 AVPTETETANSKPAKPOAQAPQTEPARENVSTNTEBPQSQ-----TATYSTEQ 1408
Qy 378 KLRIFQVVEAPAPAPPCIGGCSVAQPDVSQNSDCLIKHAATNRFLSSGKOKTKPK 437
Db 1409 PAKETSSNTEQAPREHSIMTG-----SALTW-----TETAKSDKQP 1445
Qy 438 EKVKT-----KPEKPSLPKCSYQVVGIKIS--SVHGLASEKREPVKYKWLARSBSGK 430
Db 1446 METVTENDRQPEANVTADNSVANNSSSSKSPRRRSVQPKETSAEBTTVASTOBTTVD 1505
Qy 491 EACBSSTPMASDHNYNAVKP-----EKPEKPT 519
Db 1506 NSV---STPEPNSKRTRRSVQTNSTYEPVLPPT 1534

RESULT 7

US-09-839-996-6
Sequence 6, Application US/09839996
Patent No. 6643371
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
Falkow, Stanley

TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/839,996

FILING DATE: 20-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/236,791

FILING DATE: 25-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Treccatlin, Richard P.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1848 amino acids

TYPE: amino acid

TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-839-996-6

Query Match 4.8%; Score 153.5; DB 4; Length 1848;

Best Local Similarity 21.7%; Pred. No. 0.00055;

Matches 124; Conservative 71; Mismatches 208; Indels 169; Gaps 25;

DB 8 SNEAPKAIKPTSKFRKWTGFRRTTIARREGAGDTADPSR-----QQPOQHNLRLR 61

DB 1072 STETAPKSDTATQTE-----NPNSESVPSFETEKVANPQENETVAK 1114

QY 62 SGRQ-----PKTERVEEFLTV-----RRRKNVPSLSDSEPTSTVDVE 106

DB 1115 NGEATEPTPQNGEYAKEDQPTVEANTQNEATQSEKTEETQATKSEPTES-VTVSE 1173

QY 107 TASBSVSSSSIRSGPVSLSLGEHPASSKAKGSEEBEDTSDSDGLTLKELQNRRLR 166

DB 1174 NOPEKTVSOSTE-----DKVYV--KEEKAK--VETETOKAQ--VTSKEPPKQAR 1219

QY 167 RRKEQEPVERSLRGSQNRLLKRRERDSAEYSVQISAEQ--DRPLCKQEP-- 217

DB 1220 PAPEEVPITDINABEALQQTQPTVAALFTTSPNSKPAEBETQPSBKTNABEYTVVSE 1279

QY 218 --ASGQPVQSSTDDIENLQEGKATQNTENPR--AGKP--KPECEYVDNALYC 268

DB 1280 NTATQ--PTETETATKYEKEKTEQEVPOVASQESPKQOPAKPOAQTKPOABEPAREVTL-- 1336

QY 269 ICRQPHNNRPMICCRCEMTHFGDCVIGSEARGRLLEBRGEDIYCNCTILQVODETNGS 328

DB 1337 -----TTKNVGBPO-----PQAPQPTOST 1355

QY 329 A---TNEODSGCRSVGADGTDG-----TSIGYVQKSGEDQIGKRIEKAANPSGKK 377

DB 1356 AVPTTGTAANGKPAKPAQAKPQTEPARBENVSTVNTBEPQSQ-----TSATVSTEQ 1408

QY 378 KAKTFQPVVEAFGAPKCTIGPGCSSVAQPDVSYCSNDCLKHAATTRPLSSGEGQTKPK 437

DB 1409 PAKETSSNVEOPAPENSINTG-----SATM--TETAESKDKPO 1445

QY 438 EKVKY-----KPEKPSLPRKCSYQVVGIKIS---SVHKRLASEKREMPYKVMLASRSETSGK 490

DB 1446 METVTENRQPEANVTVAANSVANNSSSESKSRBRRSVQPRBTSMEETVASTOETTYD 1505

QY 491 EAACSSSTPMSADNHNNAVKP---EKPEKPT 519

DB 1506 NSV---STPKPRSRRTSRVSQTNVSEYVPELPT 1534

RESULT 8

US-10-080-505-6

Sequence 6, Application US/10080505

Patent No. 6676948

GENERAL INFORMATION:

APPLICANT: St. Gene, Joseph W.

TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS

FILE REFERENCE: A-59941-1/RFT/DCP/DHR

CURRENT APPLICATION NUMBER: US/10/080,505

PRIOR FILING DATE: 2002-02-22

PRIOR APPLICATION NUMBER: US 08/296,791

PRIOR FILING DATE: 1994-10-25

PRIOR APPLICATION NUMBER: US 09/839,996

NUMBER OF SEQ ID NOS: 58

SOFTWARE: Patent version 3.1

SEQ ID NO: 6

LENGTH: 1848

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-10-080-505-6

Query Match 4.8%; Score 153.5; DB 4; Length 1848;

Best Local Similarity 21.7%; Pred. No. 0.00055;

Matches 124; Conservative 71; Mismatches 208; Indels 169; Gaps 25;

DB 8 SNEAPKAIKPTSKFRKWTGFRRTTIARREGAGDTADPSR-----QQPOQHNLRLR 61

DB 1072 STETAPKSDTATQTE-----NPNSESVPSFETEKVANPQENETVAK 1114

QY 62 SGRQ-----PKTERVEEFLTV-----RRRKNVPSLSDSEPTSTVDVE 106

DB 1115 NGEATEPTPQNGEYAKEDQPTVEANTQNEATQSEKTEETQATKSEPTES-VTVSE 1173

QY 107 TASBSVSSSSIRSGPVSLSLGEHPASSKAKGSEEBEDTSDSDGLTLKELQNRRLR 166

DB 1174 NOPEKTVSOSTE-----DKVYV--KEEKAK--VETETOKAQ--VTSKEPPKQAR 1219

QY 167 RRKEQEPVERSLRGSQNRLLKRRERDSAEYSVQISAEQ--DRPLCKQEP-- 217

DB 1220 PAPEEVPITDINABEALQQTQPTVAALFTTSPNSKPAEBETQPSBKTNABEYTVVSE 1279

QY 218 --ASGQPVQSSTDDIENLQEGKATQNTENPR--AGKP--KPECEYVDNALYC 268

DB 1280 NTATQ--PTETETATKYEKEKTEQEVPOVASQESPKQOPAKPOAQTKPOABEPAREVTL-- 1336

QY 269 ICRQPHNNRPMICCRCEMTHFGDCVIGSEARGRLLEBRGEDIYCNCTILQVODETNGS 328

DB 1337 -----TTKNVGBPO-----PQAPQPTOST 1355

QY 329 A---TNEODSGCRSVGADGTDG-----TSIGYVQKSGEDQIGKRIEKAANPSGKK 377

DB 1356 AVPTTGTAANGKPAKPAQAKPQTEPARBENVSTVNTBEPQSQ-----TSATVSTEQ 1408

QY 378 KAKTFQPVVEAFGAPKCTIGPGCSSVAQPDVSYCSNDCLKHAATTRPLSSGEGQTKPK 437

DB 1446 METVTENRQPEANVTVAANSVANNSSSESKSRBRRSVQPRBTSMEETVASTOETTYD 1505

QY 491 EAACSSSTPMSADNHNNAVKP---EKPEKPT 519

Db 1506 NSV---STPKPRSRTRRSVQTNSTYEPVELPT 1534

RESULT 9
PCT-US95-10661A-6

Sequence 6, Application PC/TUS9510661A

GENERAL INFORMATION:

APPLICANT: Washington University, et al.

TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSER: Flehr, Hobbach, Test, Albrighton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10661A

FILING DATE: 16-AUG-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/296,791

FILING DATE: 25-AUG-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Treccarlin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: FP-59941/RFT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1848 amino acids

TYPE: amino acid

TOPOLOGY: unknown

PCT-US95-10661A-6

Query Match 4.8%; Score 153.5; DB 5; Length 1848;

Best Local Similarity 21.7%; Pred. No. 0.00055;

Matches 124; Conservative 71; Mismatches 208; Indels 169; Gaps 25;

QY 8 SNEAPKAIKPTSKPRKRTTWGFRRTTIARBGAGDTADPSE-----QQPOQHLSLRK 61
DB 1072 STETAPKSDTATQTE-----NPNSESVPSEETREKVAENPQENETVAK 1114
QY 62 SGRQ-----PKRTEVBEFLTV-----RRGKNVAVSLSEDSSEPTSTTVDR 106
DB 1115 NEQEAETPTPQNGEVAKEQDQFVEANQTQNEATQSEKTEBTQTATTSSEPTES-VTVSE 1173
QY 107 TASEGSVSSSEIRSGPVSDSLGKEHPASSEKAKGSEBEEDTSDSDGLTLKEQLNRLK 166
DB 1174 NQPEKTQSGTE-----DKVYVE---KEEKAK---VETREFTQKAPQ---VTSKEPPEKQAK 1219
QY 167 RKREGEPEPERSLRGQNLARKRREDSAEFGSVQIGSBQ---DRPLCKQEPF----- 217
DB 1220 PAPEEVPPTDYNAEKAQALQOTPTTVAALETSPNSKPAEETQQPESEKTNABPVPVASE 1279
QY 218 --ASQGPVSQSEITDIDENLEGKATQNTENPRR---AGKF---KPECEVYDPAALYC 268
DB 1280 NTATQ-PTRETTATAVEKEKTOBQVAVASQSPKQEPAPAKQAQTKQAPARENVL--- 1336
QY 269 ICRQPHNNRFMIICDRCEBEPFGDCVGISSEARGLLERNAGDYICPNCITLLQVODETNGS 328
DB 1337 -----TTKNVGEBPQ-----PQAQPTQGST 1355

QY 329 A---TNEQDSCGRSVGADGTD-----TSIGTVBOKSGEDQGIKGRIRKANPSGKK 377
DB 1356 AVPTTGETNANSKPAKAPQAKPQTEPARENSVTNTBPSQ-----TSATYSTIQ 1408
QY 378 KLKIFQVVEAPGAPKCIIGPCSSVAQPDVYCSNDCLIKHAATNRFISGREGQTKPK 437
DB 1409 PAKETSSNVEQPAPENSINTG-----SATIM---TETAEKSDKQ 1445
QY 438 EKVKT---KPEKSLPKCSVGVGKIS---SVHKLASKEKENPVKVMASRSTSGK 490
DB 1446 METVTENDQFPANTVADNSVANNSESSSKRRRRSVQPKETSALETTVAJSTOETVVD 1505
QY 491 EAACESTPSWASDENYNAVCR---EKPEKPT 519
DB 1506 NSV---STPKPRSRTRRSVQTNSTYEPVELPT 1534

RESULT 10

US-08-285-440-5

Sequence 5, Application US/08285440

Patent No. 5532337

GENERAL INFORMATION:

APPLICANT: Ken'ichiro HAYASHI et al.

TITLE OF INVENTION: POLYPEPTIDE

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSER: Wenderoth, Lind & Ponack

STREET: 805 Fifteenth Street, N.W., #700

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 500 kb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Displaywrite

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/285,440

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/858,947

FILING DATE: March 27, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-8850

TELEFAX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 532 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL:

ANTI-SENSE:

FRAGMENT TYPE:

ORIGINAL SOURCE:

ORGANISM:

STRAIN:

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE:

HABITYPE:

TISSUE TYPE:

CELL TYPE:

CELL LINE:

ORGANELLE:

IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-5

Query Match 4.7%; Score 151.5; DB 1; Length 532;

Best Local Similarity 20.9%; Pred. No. 0.00013; Matches 120; Conservative 66; Mismatches 233; Indels 155; Gaps 23;

30 RRTTAKRGGADTEADPSQPOOHNLRLRRSGRQPK-RTREVERFLTVRRGKXNP 88
10 RRLSLALSGIAYQRNDDEBEARER---RRARQERLRQOEBSLGQVTDVAVNAQ 65
89 VSLBDSSEPTSTVTDVETASBGSVSSSEIRSGPVSDSLGKEHPASSKAKGSEBEDT 148
66 NSVPEBEAKTTTNTNOVE-----GDDEAFLRLRLARRRERQ- 102
149 SPSDSGLTLKLNRLRKREQBPVERSIRGSONRLRKREEDSAFGSVQISABED 208
103 -----KRLQALERQKEFDP---TTDASLSLPSRRQNDTAEN---ETTERE 144
209 RPLCKQBPASQSPVSQSET-----DIEHOLEGKATOGTEBNPREAG----- 252
145 E---KSEBRQERYRIBETETVTKSYQKNDWRDAENKEDKEKEBEBEKPKRGSIGENQ 201
253 -----KPKPE-----CEVYDPNALYCTICRQPH-NNRFMICCDRCBEMFH 290
202 IKDEKIKDKKPKKEHVKSFMORRKGFTBVKSGNGEFMTKLTHTBNTFSRPGGRAS--- 257
291 GDCVGISERAG-----RLERNGE-----DYICPCTIIQVODETN 326
258 ---VTKRABGAPQVEAGRLLELRRRGETSEBEFKLKQKQOEALALEBELKCKRER 314
327 GSATNBDSGCSVAGDGTCTSIGTVEQKSGEDQIKRIRKAAAPSGKKLKIFQPVV 386
315 RKLVEERQRRKQEBAD-----RKLAREBEKRLKEIIRRRARAAEKQKQKPEBDL 366
387 BAPGAP-KCIGPGCSVAQPSVCSNDCILK-----HAAATMFLSSGEOKTKPRE 438
367 SDDKPFKCFPTPKGSLLKIBERAEFLANKSVQKSGVSKTHQAIAIVSKIDSRLEQYTSAL 426
439 KVK-----TKRPFSLPKCSVQVIGIKIISVHRLASBKKEPNVKTWMLAS--RSETSGKEA 492
427 GTSKAKPTKPAASDLP-----VPAEGVRIKSNMKEKGNVPSPTAAGTNPETAGLKV 479
493 ACESSTPSMAS---DHNTAAVAPKPEKPTALSP 523
480 GVSSRIINEMUTPTDGN-----KSPAP-KPSDLRP 508

RESULT 11
US-08-630-349-5

Sequence 5, Application US/08630349
Patent No. 5739008
GENERAL INFORMATION:
APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,349
FILING DATE: April 10, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/285,440
FILING DATE: August 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren W. Cheek, Jr.
REGISTRATION NUMBER: 33,367
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-5

Query Match 4.7%; Score 151.5; DB 1; Length 532;

Best Local Similarity 20.9%; Pred. No. 0.00013; Matches 120; Conservative 66; Mismatches 233; Indels 155; Gaps 23;

30 RRTTAKRGGADTEADPSQPOOHNLRLRRSGRQPK-RTREVERFLTVRRGKXNP 88
10 RRLSLALSGIAYQRNDDEBEARER---RRARQERLRQOEBSLGQVTDVAVNAQ 65
89 VSLBDSSEPTSTVTDVETASBGSVSSSEIRSGPVSDSLGKEHPASSKAKGSEBEDT 148
66 NSVPEBEAKTTTNTNOVE-----GDDEAFLRLRLARRRERQ- 102
149 SPSDSGLTLKLNRLRKREQBPVERSIRGSONRLRKREEDSAFGSVQISABED 208
103 -----KRLQALERQKEFDP---TTDASLSLPSRRQNDTAEN---ETTERE 144
209 RPLCKQBPASQSPVSQSET-----DIEHOLEGKATOGTEBNPREAG----- 252
145 E---KSEBRQERYRIBETETVTKSYQKNDWRDAENKEDKEKEBEBEKPKRGSIGENQ 201
253 -----KPKPE-----CEVYDPNALYCTICRQPH-NNRFMICCDRCBEMFH 290
202 IKDEKIKDKKPKKEHVKSFMORRKGFTBVKSGNGEFMTKLTHTBNTFSRPGGRAS--- 257
291 GDCVGISERAG-----RLERNGE-----DYICPCTIIQVODETN 326
258 ---VTKRABGAPQVEAGRLLELRRRGETSEBEFKLKQKQOEALALEBELKCKRER 314
327 GSATNBDSGCSVAGDGTCTSIGTVEQKSGEDQIKRIRKAAAPSGKKLKIFQPVV 386
315 RKLVEERQRRKQEBAD-----RKLAREBEKRLKEIIRRRARAAEKQKQKPEBDL 366
387 BAPGAP-KCIGPGCSVAQPSVCSNDCILK-----HAAATMFLSSGEOKTKPRE 438

```

Db      367 SDDKPFKCFKPKSSSLKIEERABFLNKSVGSKSTHQAALVSKIDSLRQYTSALR 426
Qy      439 KVK-----TKPEKSLPKSVGVGICISVHKRLABKRNPKYKMLAS--RSEISGKGA 492
Db      427 GTKAKAPKPKASDLR-----VPAEGVRNIKSMWKGWVFSSPTAAGTGNKETAALKV 479
Qy      493 ACESSTPSMAS--DHNYNAVPEKPEKPTALSP 523
Db      480 GVSSRINEMLTKTPDGN-----KSPAP-KPSDLRP 508

```

RESULT 12

```

US-09-976-594-726
; Sequence 726, Application US/0976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

```

```

Query Match      4.6%; Score 149.5; DB 4; Length 2468;
Best Local Similarity 20.5%; Pred. No. 0.0018;
Matches 134; Conservative 87; Mismatches 250; Indels 183; Gaps 31;

```

```

Qy      7 LSNBAPALKFT-----SKFRKTMGRKRTT---IAKREGADTBADPSEQPOQHNLIS 58
Db      861 IEDBEKLKETEPEVAVVIOKEREVTKGPAESPDEGTTTBEGGBCEQPEBELRP----- 914
Qy      59 LBSRGQRKRTREVAEPLTTVRRKQNPVSLSDSEPTSTVTVVETASGVSSSSSE 118
Db      915 VEKQG-----VDIERK-----EDGAGPSSSSS-----TGDIYBKAE 947
Qy      119 IRSGVSDSLGKEHPASSSEKANGGEEDTSDSDGLTKELQNLKREKRE----- 170
Db      948 TEAEPEDEDEHCVASAKHSPTEDBSAKAEADAY-----IRKRSVASGDOR 999
Qy      171 -QEPVRSGL-RGSQNLKRRKREDSAEYG-----SVQIGSAE 206
Db      1000 AEDMDRAIKKEGAQSEBEADBEKADABAREEYEPKMEADYVMVAVDAAEKGAE 1059
Qy      207 QD-----RPLCQ-----PEAS-----QGPVSOSTDDIENQ-----LEGRATQGN 243
Db      1060 EGYGLTTPYKQLAQSFGREPASSIHDETLFGSGSESEATASDEENREDQPEEFTATSGY 1119
Qy      244 TEENPRBAGKPECEVVDNALYCICRQPHNN-----RFMICCDRCSEWTFHGD----- 292
Db      1120 TOSTIBISSEPTPMDEMSTPRD---VMSDETNNBETESPQGFVNTTKYESSLYGQERYK 1176
Qy      293 -----CVGISARGLLERNGEDYICPNCTIL--QVDETNGSATNEODSGCRSYGADG 344
Db      1177 PADVPLNGFSEHG-EXTATDQDYNAASATISPPSSMEEDKFSRSALADAYCSEYKAST 1235
Qy      345 T-----DCSISGIVEKSGSGEDGIGKRIKAAVPSGKKLKIFQPVYVBARAGAKCIGPGCS 400
Db      1236 TLIDIOSISAVSSSEKVSPS---KSPSLSPPSPPLEKTPLEGRSVNLSLTNEILVNSAE 1291
Qy      401 SVAGPDSVYVCSNDCILKHAAL---TMRPLS-----SGKEQKTK--PRE 438

```

```

Db      1232 AEVAVSEVAVQVEVEEHCAPEDEKTLLEVSPSGVTSAGHTPYQSPTEKSSHLPTK 1351
Qy      439 KVKTEPEKPSLPKCSVOYGIRISV--HKRLAS-----EKRENPKVNLASRS-ET 487
Db      1352 VIKRP-----AVPSFEPSDAKDENERASVSPMBPVPDSDSPLEKVLSPLRSPPL 1403
Qy      468 SGKEACSES--STPSMASDHNYNAVPEKPEK--PTALSP--LISKTTTPK 534
Db      1404 IGSBSAYSEFLSADDKASGAGSESPFEKSGKQSPDQVPSVSEMTSTSLYODK 1457

```

RESULT 13

```

US-08-893-852A-3
; Sequence 3, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 53041
US-08-893-852A-3

```

```

Query Match      4.6%; Score 148; DB 3; Length 657;
Best Local Similarity 19.4%; Pred. No. 0.00036;
Matches 128; Conservative 64; Mismatches 241; Indels 226; Gaps 26;

```

```

Qy      41 GPTADPSEQPO-QHNLSLRSGQPKRT---BRVERP-----LTTVRRGKX 85
Db      75 GHTESGSEPSQAARLCLVAESSPPTWGLSNVDENAPGQODDLREKMERIAGTA 134
Qy      86 NV-PVSLDESEPTSTVTVVETASGVSSSEIRSGPVSDS----- 127
Db      135 TLQPAQLQADRLAEVAVAREGVAPAY-PTSQLEGGFAENBDEGTATKYQASASTA 193
Qy      128 -----LGR-EHPASSSEKANGGEEDTSDSDGLTKELQNLKRRKREO--- 172

```


Db 194 PGYKSTVPVPLGAEHQATBE--KGTENKADPSNSPSSGSHSRAMEVYSREKPKQEGEA 251

Qy 173 PVERSLRGSQNLARKKREEDSAETGVSQIG-----SABOD 208

Db 252 KVEARAGQGHPCRNABABEGGPETTFVCTGNAFLKAWYRRPGEDTEBEDNSDSABED 311

Qy 209 RPLCKQEPAS-----OGPVOSSETDUIENOLEGATQANTBENPREAGKPEPCE 259

Db 312 TAQGTATHTSAFLKAWYRRPGEDTEBEDSDSAEDTAQTGAT----- 356

Qy 260 VYDPNALYICQPHNNRPMICCDRCBWFHGDVGISBARGRLERNGEDYICPNCTIL 319

Db 357 -----PHTSAFL-----KAWVY-----RPGEDTEBEDSLD 382

Qy 320 QVQDETNGSATNEODSGCRS--VGADGTCTSIGTVEOKSGEDQIGKRIEKAANSKGK 377

Db 383 SAEEDTAQTGATPHTSAFLKAWYRRPGED-----TEENSDLSABEDTAQTGATPHTSP 437

Qy 378 KLK--IFQP-----VVEAPGAPKCIGPCSSVAQPDVYCSNDCILKHAATMRFLSSG 429

Db 438 FLKAWYRRPGEDTEBEDTEBEDSENVAPGDSSTASOSFCL----- 479

Qy 430 KEQKTKPEKVK-----TKPEK--PSLPKCSV----- 454

Db 480 QPQRLCPGKTKGRGBBPLFOVAFTLPGEKDESPWAAPKFLRLQRLRLPKAPTRDQD 539

Qy 455 -QVGKISSVH-----KRLASEKRENPKVY-----MLASRSETSGKE 491

Db 540 PEIPLKARVHFAKRYVHFLAVMAGPAARGPHEQFADRSRPARIRIAQABEKLGPY 599

Qy 492 AACSESTPSMASDHYNAVKPEKPEKPTALSPILLSKCTYHP-KAGFPGPSHHLGCLG 549

Db 600 LTPDSRAAMARLIRNPSLPQSE-PRSSSEATPLTDVTPPLPSETPPSPSLYLGRRG 657

RESULT 14

US-08-821-818-3

; Sequence 3, Application US/08821818

; Patent No. 616877

; GENERAL INFORMATION:

; APPLICANT: Fisher, Paul B.

; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION

; TITLE OF INVENTION: ELLEVATED GENE-3 AND USES THEREOF

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/821,818

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Chan, Albert Wai-Kit

; REGISTRATION NUMBER: 36,479

; REFERENCE/DOCKET NUMBER: 51523

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400

; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 657 amino acids

; TYPE: amino acid

; STRANDBENESS: single

; TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-821-818-3

Query Match 4.6%; Score 148; DB 3; Length 657;

Best Local Similarity 19.4%; Pred. No. 0.0006;

Matches 128; Conservative 64; Mismatches 241; Indels 226; Gaps 26;

Qy 41 GDTADPSEQPPQ-QHNLSLRSGROPKRT---ERVEEP-----LTTVRRGKK 85

Db 75 GETESGSPBQQAQRCLVAESPPETWGLSNDEYNAPRGODLREKMERATAGKA 134

Qy 86 NV-PVSLDESEPTSTYTDVETASRGVSSESEIRGVSDS----- 127

Db 135 TLQPAQLAGADRRLGEEVAVREGVAEPAY-PTSOLEGGAEEDDEITVYQASAA51A 193

Qy 128 -----LGR--RHPSSEKAVGEEHEPDSQGLTIKXONLRRRBOE--- 172

Db 194 PGYKSTVPVPLGAEHQATBE--KGTENKADPSNSPSSGSHSRAMEVYSREKPKQEGEA 251

Qy 173 PVERSLRGSQNLARKKREEDSAETGVSQIG-----SABOD 208

Db 252 KVEARAGQGHPCRNABABEGGPETTFVCTGNAFLKAWYRRPGEDTEBEDNSDSABED 311

Qy 209 RPLCKQEPAS-----OGPVOSSETDUIENOLEGATQANTBENPREAGKPEPCE 259

Db 312 TAQGTATHTSAFLKAWYRRPGEDTEBEDSDSAEDTAQTGAT----- 356

Qy 260 VYDPNALYICQPHNNRPMICCDRCBWFHGDVGISBARGRLERNGEDYICPNCTIL 319

Db 357 -----PHTSAFL-----KAWVY-----RPGEDTEBEDSLD 382

Qy 320 QVQDETNGSATNEODSGCRS--VGADGTCTSIGTVEOKSGEDQIGKRIEKAANSKGK 377

Db 383 SAEEDTAQTGATPHTSAFLKAWYRRPGED-----TEENSDLSABEDTAQTGATPHTSP 437

Qy 378 KLK--IFQP-----VVEAPGAPKCIGPCSSVAQPDVYCSNDCILKHAATMRFLSSG 429

Db 438 FLKAWYRRPGEDTEBEDTEBEDSENVAPGDSSTASOSFCL----- 479

Qy 430 KEQKTKPEKVK-----TKPEK--PSLPKCSV----- 454

Db 480 QPQRLCPGKTKGRGBBPLFOVAFTLPGEKDESPWAAPKFLRLQRLRLPKAPTRDQD 539

Qy 455 -QVGKISSVH-----KRLASEKRENPKVY-----MLASRSETSGKE 491

Db 540 PEIPLKARVHFAKRYVHFLAVMAGPAARGPHEQFADRSRPARIRIAQABEKLGPY 599

Qy 492 AACSESTPSMASDHYNAVKPEKPEKPTALSPILLSKCTYHP-KAGFPGPSHHLGCLG 549

Db 600 LTPDSRAAMARLIRNPSLPQSE-PRSSSEATPLTDVTPPLPSETPPSPSLYLGRRG 657

RESULT 15

US-09-052-753B-3

; Sequence 3, Application US/09052753B

; Patent No. 6472520

; GENERAL INFORMATION:

; APPLICANT: Paul B. Fisher

; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses

; TITLE OF INVENTION: Thereof

; FILE REFERENCE: A34608-B

; CURRENT APPLICATION NUMBER: US/09/052,753B

; CURRENT FILING DATE: 1998-03-31

; PRIOR APPLICATION NUMBER: PCT/US98/05793

; PRIOR FILING DATE: 1998-03-20

; PRIOR APPLICATION NUMBER: 08/812,818

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 657

; TYPE: PRT

; ORGANISM: rat

PS Claim 8; Fig 1D; 27pp; English.

XX The invention provides nucleic acids encoding the human and murine death
CC inducer-oblierator 1 (DIO-1) polypeptides. The polypeptides can be
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC agonists and antagonists are used as a medicament for treating diseases
CC characterized by an alteration in cell death or by hyperproliferation,
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC tumours, malignant tumours or hyperproliferative skin disorders. They are
CC also useful in the treatment of metabolic, proliferative or inflammatory
CC conditions. The present sequence represents the murine DIO-1 polypeptide
XX
SQ Sequence 614 AA;

Query Match 100.0%; Score 3228; DB 3; Length 614;
Best Local Similarity 100.0%; Pred. No. 1,4e-215;
Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MODKGLSNEBAKAIKPTSKERKRWGFRRTTIARBGAGTDEADPSEQPOQHNLIR 60
DB 1 MODKGLSNEBAKAIKPTSKERKRWGFRRTTIARBGAGTDEADPSEQPOQHNLIR 60
QY 61 RSGRQPKRTREVEFLITVRRRGGKNVPVSLDSEPTSTVTVDVETASGVSSSSIR 120
DB 61 RSGRQPKRTREVEFLITVRRRGGKNVPVSLDSEPTSTVTVDVETASGVSSSSIR 120
QY 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLTKELQNLRLRKROEPEVRSILG 180
DB 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLTKELQNLRLRKROEPEVRSILG 180
QY 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLTKELQNLRLRKROEPEVRSILG 180
DB 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLTKELQNLRLRKROEPEVRSILG 180
QY 181 SGNRLRKRRREDSATSGVQISARQDRPLCKOEPEASOGPVSQSEETDIEQLSGKAT 240
DB 181 SGNRLRKRRREDSATSGVQISARQDRPLCKOEPEASOGPVSQSEETDIEQLSGKAT 240
QY 241 QGNTENPREAKPKPECEVYPNALYICICROPNNRPMICDRCCEWTFHGDVCGISBAR 300
DB 241 QGNTENPREAKPKPECEVYPNALYICICROPNNRPMICDRCCEWTFHGDVCGISBAR 300
QY 301 GLLERNGEDYICPNCTIIQVODETNGSATNEBDSGCRSGVADGTCTSIGTVEQSGED 360
DB 301 GLLERNGEDYICPNCTIIQVODETNGSATNEBDSGCRSGVADGTCTSIGTVEQSGED 360
QY 361 GQIKRIRKAAVPSGKKIKITPPVEAGAPKICPGSSVAPQSPVSCSNDCLIKHAA 420
DB 361 GQIKRIRKAAVPSGKKIKITPPVEAGAPKICPGSSVAPQSPVSCSNDCLIKHAA 420
QY 421 ATMRFLSSGKEQKTKPEKVKTKPEKPSLPKCSVOVGIIKISVHKRLASKEKNPVKRW 480
DB 421 ATMRFLSSGKEQKTKPEKVKTKPEKPSLPKCSVOVGIIKISVHKRLASKEKNPVKRW 480
QY 481 LASRSETSGKEAACSSTPSWASDHNNYNAVKPEKPEKPTALSTTLISKCTYHKAAPGP 540
DB 481 LASRSETSGKEAACSSTPSWASDHNNYNAVKPEKPEKPTALSTTLISKCTYHKAAPGP 540
QY 541 SHHIGGCLGLSTRVYGVLTIVASSSLPARSRYODASGQVPLPSLMSLSGFTLSGCV 600
DB 541 SHHIGGCLGLSTRVYGVLTIVASSSLPARSRYODASGQVPLPSLMSLSGFTLSGCV 600
QY 601 LMLBAISYSPRPW 614
DB 601 LMLBAISYSPRPW 614

```

RESULT 2
AAB93638
ID AAB93638 standard; protein; 562 AA.

XX AAB93638;
AC
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:13130.
XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX BP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000BP-00116126.

XX 29-JUL-1999; 99QP-00248036.

XX 27-AUG-1999; 99QP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 13130; 2537pp + Sequence listing; English.

PS The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 562 AA;

Query Match 65.4%; Score 2110; DB 4; Length 562;
Best Local Similarity 75.0%; Pred. No. 5.2e-138;
Matches 420; Conservative 32; Mismatches 94; Indels 14; Gaps 7;

```

QY 1 MODKGLSNEBAKAIKPTSKERKRWGFRRTTIARBGAGTDEADPSEQPOQHNLIR 58
DB 1 MODKGLSNEBAKAIKPTSKERKRWGFRRTTIARBGAGTDEADPSEQPOQHNLIR 60
QY 59 LRSRQPKRTREVEFLITVRRRGGKNVPVSLDSEPTSTVTVDVETASGVSSSSIR 118
DB 61 LRSRQPKRTREVEFLITVRRRGGKNVPVSLDSEPTSTVTVDVETASGVSSSSIR 120
QY 119 IRSGPVSDSLG-KEHPASSEKAKGSEEBTSDSDGLTKELQNLRLRKROEPEVRS 177
DB 121 TRSGPOSASTAVAKERASSEKAKGSDHDDTSDSDGLTKELQNLRLRKROEPEVRS 180
QY 178 LRSQNLRLRKRRREDSATSGVQISARQDRPLCKOEPEASOGPVSQSEETDIEQLSG 236

```

Db 181 LKGIQSRLRKKRREBPAETVGSASDVBGLP-SKQEPENDQGVVSQAGKDRESKLE 239

Qy 237 GKATQGTENPRPRAKPKPCPCRYDPAALYCICQPPNNRMICCRCEBPHGDCVI 296

Db 240 GKAAQDIDKEBPDIGRPCEBEGYDPNALYCICQPPNNRMICCRCEBPHGDCVI 299

Qy 297 SEARGRLLENGEDYICPNCCTILQVODETNGSATNBODSGRSVAGDGTCTSIGTVEOK 356

Db 300 SEARGRLLENGEDYICPNCCTILQVODETHSETDAQAKRRPGDADDTCTSIGTIEOK 359

Qy 357 SGEODGIGRIEKAANPSGKKKTKIPOPVBARPAKPCIGPCSSVAOPDSVYCSNDCIL 416

Db 360 SSEDGIGRIEKAANPSGKKKTKIPOPVBARPAKPCIGPCSSVAOPDSVYCSNDCIL 419

Qy 417 KHAATMRFLSSGKEQTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERRENPV 476

Db 420 KHAATMRFLSSGKEQTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERRENPV 479

Qy 477 KK-VMLASRSETSGKBAACBSSTPSMASDHNYNAYKPEKPEPTALSPTLISKCTYHPKA 535

Db 480 KKAVVVPARSRBALGKBAACBSSTPSMASDHNYNAYKPEKPTAP---SPSLAYKCMYHLGV 536

Qy 536 GPPGPSHHL-----GGCLGL 550

Db 537 GLDPSRSFWAIAPWACPGL 556

RESULT 3

AAV67579 standard; protein; 562 AA.

AAV67579:

AAV67579:

19-JUN-2000 (first entry)

Human death inducer-obliterator 1 (DIO-1) polypeptide.

Death inducer-obliterator 1: DIO-1; cell death; cancer; tumour; human; autoimmune disease; cytostatic; immunosuppressive; antidiabetic; antineutritic; antiinflammatory; antiproliferative.

Homo sapiens.

Key Location/Qualifiers

Region 165..172

Region /note= "NLS sequence"

Region 185..193

Region /note= "NLS sequence"

Region 271..288

Region /note= "zinc finger motif"

Region 293..320

Region /note= "zinc finger motif"

MO200015787-A1.

23-MAR-2000.

10-SEP-1999; 99MO-GB003019.

10-SEP-1998; 98SR-00003069.

17-SEP-1998; 98US-0100873P.

(CMSJ) CONSELTO SUPERIOR INVESTIGACIONES CIENTIF.

(BANN/) BANNERMAN D G.

Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;

WPI: 2000-271426/23.

N-PSDB; AA290578.

New DNA encoding human and murine death inducer-obliterator 1 polypeptides, useful in the treatment of cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumors, malignant tumors and

PT hyperproliferative skin disorders.

PS Claim 6, Fig 1C; 27pp; English.

CC The invention provides nucleic acids encoding the human and murine death inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be expressed by standard recombinant methodology. The DIO-1 polypeptides, agonists and antagonists are used as a medicament for treating diseases characterized by an alteration in cell death or by hyperproliferation, e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumors, malignant tumors or hyperproliferative skin disorders. They are also useful in the treatment of metabolic, proliferative or inflammatory conditions. The present sequence represents the human DIO-1 polypeptide

Sequence 562 AA:

Query Match 65.3%; Score 2108; DB 3; Length 562;

Beat Local Similarity 75.0%; Fred. No. 7.1e-136;

Matches 420; Conservative 32; Mismatches 94; Indels 14; Gaps 7;

Qy 1 MDDKGLSNEBAPRAIKPTSKPFRKTGFRRTTIARBAAGPTDAPSEQCP--OOHNLIS 58

Db 1 MDDKGDPSNEBAPRAIKPTSKPFRKTGFRRTTIARBAAGADPAPLEPPPOOGLIS 60

Qy 59 LRSRGQPKRTERYEBPLTVARRGKNVPSVLADSSSEPTSTVTVDVTASGVSSESSB 118

Db 61 LRSRGQPKRTERYVQFLTIARRGRSMVSLSDSGSEPTSCPATDAETASGVSASB 120

Qy 119 TRSGVSPSLG-KHPASSERAKGSEBBDTSDDGLTKELQNRRLARKRQERVERS 177

Db 121 TRSGQASATVAKERPASSERAKGDDDDTSDSDGLTKELQNRRLARKRQERTERP 180

Qy 178 LRSGQNRRLARKRREDSABT-GSVQIGSABQDRPLCKQSPBASQSPVSGSFTDLENGLE 236

Db 181 LKGIQSRLRKKRREBPAETVGSASDVBGLP-SKQEPENDQGVVSQAGKDRESKLE 239

Qy 237 GKATQGTENPRPRAKPKPCPCRYDPAALYCICQPPNNRMICCRCEBPHGDCVI 296

Db 240 GKAAQDIDKEBPDIGRPCEBEGYDPNALYCICQPPNNRMICCRCEBPHGDCVI 299

Qy 297 SEARGRLLENGEDYICPNCCTILQVODETNGSATNBODSGRSVAGDGTCTSIGTVEOK 356

Db 300 SEARGRLLENGEDYICPNCCTILQVODETHSETDAQAKRRPGDADDTCTSIGTIEOK 359

Qy 357 SGEODGIGRIEKAANPSGKKKTKIPOPVBARPAKPCIGPCSSVAOPDSVYCSNDCIL 416

Db 360 SSEDGIGRIEKAANPSGKKKTKIPOPVBARPAKPCIGPCSSVAOPDSVYCSNDCIL 419

Qy 417 KHAATMRFLSSGKEQTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERRENPV 476

Db 420 KHAATMRFLSSGKEQTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERRENPV 479

Qy 477 KK-VMLASRSETSGKBAACBSSTPSMASDHNYNAYKPEKPEPTALSPTLISKCTYHPKA 535

Db 480 KKAVVVPARSRBALGKBAACBSSTPSMASDHNYNAYKPEKPTAP---SPSLAYKCMYHLGV 536

Qy 536 GPPGPSHHL-----GGCLGL 550

Db 537 GLDPSRSFWAIAPWACPGL 556

RESULT 4

AAV67579 standard; protein; 1191 AA.

AAV67579:

AAV67579:

06-NOV-2001 (first entry)

Human protein SBQ ID NO 3865.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

tissue growth factor; immunomodulatory; cancer; leukaemia;
 nervous system disorder; arthritis; inflammation.
 Homo sapiens.
 MO200157190-A2.
 09-AUG-2001.
 05-FEB-2001; 2001MO-US004098.
 03-FEB-2000; 2000US-00496914.
 27-APR-2000; 2000US-00560875.
 20-JUN-2000; 2000US-00598075.
 19-JUL-2000; 2000US-00620325.
 01-SEP-2000; 2000US-00654936.
 15-SEP-2000; 2000US-00663561.
 20-OCT-2000; 2000US-00693325.
 30-NOV-2000; 2000US-00728422.
 (HYSB-) HYSEQ INC.
 Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZM; Xue AJ, Yang Y, Wejhrman T, Goodrich R; WPI; 2001-476283/51.
 N-PSDB; AAK53352.
 Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.
 Claim 20; Page 455-456; 6221pp; English.
 The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoietic regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication
 Sequence 1191 AA:
 Query Match 55.5%; Score 1792.5; DB 4; Length 1191;
 Best Local Similarity 70.6%; Pred. No. 1.5e-115;
 Matches 358; Conservative 32; Mismatches 74; Indels 43; Gaps 6;

361 OGKGRIEKAAPSXXXXKIFQP----- 384
 |||||
 321 OGKGRIEKAAPSXXXXKIFQPGPVPVQLPVLQVLEIAVSRKSIAPFLHCTISCK 380
 |||||
 385 VVEAPGAPKCTIGPGSSVAQPDVYCSNDCTIKHAATATPLSGKEQTKREKXTKP 444
 |||||
 381 VTEAPGAKCTIGPGCCHVAQPDVYCSNDCTIKHAATATPLSGKEQTKREKXTKP 440
 |||||
 445 EEPSPKCSVQVIGITSSVHRKLAASEKENPVKK-VMLASRETSGKAAECSSSTPSMAS 503
 |||||
 441 EEPSPKCSVQVIGITSSVHRKLAASEKENPVKK-VMLASRETSGKAAECSSSTPSMAS 500
 |||||
 504 DHNYNAVKEPEKEPTALSPFLTSKCT 530
 |||||
 501 DHNYNAVKEPEKTAAP---SPSLTYKST 524
 |||||
 RESULT 5
 ABG22389
 ID ABG22389 standard; protein; 775 AA.
 XX
 AC ABG22389;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #22380.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001MO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSB-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS86576.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 52748; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as a hybridisation probe, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this
 CC patient did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 775 AA;

Query Match 54.4%; Score 1756.5; DB 4; Length 775;
 Best Local Similarity 57.1%; Pred. No. 2.7e-113;
 Matches 378; Conservative 45; Mismatches 126; Indels 113; Gaps 14;

DB 1 MODKHLNNEBAPKAIKPTSKPRKTWGRRTTIANGEGAGTADPPSEQP--QGHNIS 58
 DB 131 MODKDPNNEBAPKAIKPTSKPRKTWGRRTTIANGEGAGTADPPSEQPQOOLGLS 190
 QY LRSRGRPRTRRVEREFTTVRRGKKNPVSLDESBPTSSVTVTETASGCVSSSS 118
 DB 191 LRSRGRPRTRRVEREFTTVRRGKKNPVSLDESBPTSSVTVTETASGCVSSSS 250
 QY LRSRGRPRTRRVEREFTTVRRGKKNPVSLDESBPTSSVTVTETASGCVSSSS 177
 DB 251 TRSGFQASATVAKERPAKSEKTKGDDHDDTSDSDGLTKELQNLRLRKREQEVERS 310
 QY 178 LRSQNLRLRKREBDSAEFT--GSVQIGABDRPLCKQEPKASQGVSGSETVDIENQL 236
 DB 311 LKIGSRLLKRRKREBDSAEFT--GSVQIGABDRPLCKQEPKASQGVSGSETVDIENQL 369
 QY 237 GATVGNTEANREACKPRCEVDPNLYCICRPHNNRPMICCDRCBEMFHGCVGI 296
 DB 370 GATVGNTEANREACKPRCEVDPNLYCICRPHNNRPMICCDRCBEMFHGCVGI 429
 QY 297 SEARGLTERNGEDYICPNCTILQVDETNAGATNEDSGCSVVG--ADGTDCTSIGTVEQ 355
 DB 430 SEARGLTERNGEDYICPNCTILQVDETNAGATNEDSGCSVVG--ADGTDCTSIGTVEQ 489
 QY 356 K--SGBDGQIKRIRERKANPSGKKKLTQF-----VTEAPGA 391
 DB 490 RELSRPRPKLKRIRERKANPSGKKKLTQF-----VTEAPGA 549
 QY 392 PKCIPRGCSVAVOPVSGNCTILKHAATWRFSSGKQKPKKPKKPKKPKKPKKPKK 451
 DB 550 PKCIPRGCSVAVOPVSGNCTILKHAATWRFSSGKQKPKKPKKPKKPKKPKKPKK 609
 QY 452 C-----SVQVG-----IKISSVHKLASKEKREPNPKVW 480
 DB 610 CAAQVYQVMTLLSSWQLGSSGLQVTLCPPTTHNMLKFSVLHGTAQPPRHRIRKL- 668
 QY 481 LASRSETSGKBAACBSSTPMSADHNNYNAVKEPKPEPTALSPILLKCTYHPKAGPPGP 540
 DB 669 -----NKYMGACHCMPLRP-----LPCL-----PCMPRS 693
 QY 541 SHHLGGCGLSTRV-----LGVVLVIVASSSLPARBRVQDASGPVYPLBSLMSGW 593
 DB 694 SPSLVGAGSLPRHAAVHLNPPCLAI-----CPLAARLSPARGLGSDDLTWGYKWW 744
 QY 594 FL 595
 DB 745 LL 746

RESULT 6
 AAM79235
 ID AAM79235 standard; protein; 647 AA.

AC AAM79235;

DT 06-NOV-2001 (first entry)

DE Human protein SBQ ID NO 1897.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;

KX nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001, 2001WO-US004098.

XX 03-FEB-2000, 2000US-00496914.

XX 27-APR-2000, 2000US-00560875.

XX 20-JUN-2000, 2000US-00598075.

XX 19-JUL-2000, 2000US-00620325.

XX 01-SEP-2000, 2000US-00654936.

XX 15-SEP-2000, 2000US-00663561.

XX 20-OCT-2000, 2000US-00693325.

XX 30-NOV-2000, 2000US-00728422.

XX (HYSR-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhtman T, Goodrich R;
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK52368.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.

XX Claim 20; Page 4293-4294; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SBQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM0020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX Sequence 647 AA;

Query Match 52.5%; Score 1693.5; DB 4; Length 647;
 Best Local Similarity 70.4%; Pred. No. 5.1e-109;
 Matches 340; Conservative 28; Mismatches 72; Indels 43; Gaps 6;

QY 87 VPVSLDESBPTSSVTVDVETASGCVSSSRTIRSGPVSDSLG--KHPASSBRKAGGBR 145
 DB 1 MPVSLDESBPTSSCPATDAETASGCVSSSASTRSGPQASATVAKKPASSBRKAGGBR 60

QY 146 ERTSDSDGLTKELQNLRLRRRBOEPVERSLRGSQNLRRKRRRDSAEFT--GSVQIGS 204
 DB 61 DRTSDSDGLTKELQNLRLRRRBOEPVERSLRGSQNLRRKRRRDSAEFT--GSVQIGS 120

QY 205 AEQDRPLCKQEPKASQGVSGSETDIEHQLEKATQANTENPREAKPRCEVDPN 264
 DB 121 VEGVLP--SKQEPENDQGVVSGAKODRSGTGBKAAQDKDBRPGDLGRPKPCGEGDPN 179

QY 265 ALYCTCRPHNNRPMICCDRCBEMFHGDCVGISBAAGLLRERGGDYICNCTILQVQDE 324
 DB 180 ALYCTCRPHNNRPMICCDRCBEMFHGDCVGISBAAGLLRERGGDYICNCTILQVQDE 239

QY 325 TNGSATNBODSGGASVAGADGTCTSIGTVKSGGEQGIKGRIRKANPSGKKKLTQF 384
 DB 240 THTSATNBODGATWRFEDADGTCTSIGTVKSGGEQGIKGRIRKANPSGKKKLTQF 299

```

Oy 365 -----VVEAPGAPKCIIGPCSSVAOPDSV 408
Db 300 GPGVPYQIPLVMQVLEIAVRSISAPFTLLHICISCVIAPASAKICIGGCGHVAOPDSV 359
Oy 409 YCSNDICILKHAATNRPLSSGKQKTPKPKVKTPEKESLPKCSQVQVGIKISVHKRLA 468
Db 360 YCSNDICILKHAATNRPLSSGKQKTPKPKVKTPEKESLPKCSQVQVGIKISVHKRLA 419
Oy 469 SEKRNPNYK-VMLASRSTSGKEAACESTPSMASDHNHNAVKPEKPEPTALSPPTLS 527
Db 420 PEKKTTPYKAVVAPARSEALGKEAACESTPSMASDHNHNAVKPEKTAAP---SPSLLY 476
Oy 528 KCT 530
Db 477 KST 479

RESULT 7
ABG22388
ID ABG22388 standard; protein; 330 AA.
XX
AC ABG22388;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22379.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PI 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-MAR-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
DR N-PSDB; AAS86575.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostic, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 52747; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantifying a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostic, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this

```

```

CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 330 AA;
XX
Query Match 37.9%; Score 1224.5; DB 4; Length 330;
Best Local Similarity 70.7%; Pred. No. 7.9e-77;
Matches 232; Conservative 21; Mismatches 36; Indels 37; Gaps 2;
XX
Oy 232 ENQLRGKATQGTENPNRPAKPKPCBYVDNALYCTIRQPHNNRPMICCRCEWTFH 291
Db 1 ESKLKGKAAQDIKDESPGDIAPKPCBCGCDYDPAALYCIQPHNNRPMICCRCEWTFH 60
Oy 292 DCVGISSEARGRLLENGRDYICPNCTIIQVODBTNGSATNEODSGCRVAGADTDCSTIG 351
Db 61 DCVGISSEARGRLLENGRDYICPNCTIIQVODBTNGSATNEODSGCRVAGADTDCSTIG 120
Oy 352 TVEOKSGEDQIGKRIEKANPSGKKQLKIFQF----- 384
Db 121 TIEQKSSBDQIGKRIEKANPSGKKQLKIFQFPGPVPYQIPLVMQVLEIAVRSISAP 180
Oy 385 -----VVEAPGAPKCIIGPCSSVAOPDSVYCSNDICILKHAATNRPLSSGKQKTPK 435
Db 181 TLLHICISCVIAPASAKICIGGCGHVAOPDSVYCSNDICILKHAATNRPLSSGKQKTPK 240
Oy 436 PEKRVKTPKPKESLPKCSQVQVGIKISVHKRLASEKRNPNYK-VMLASRSTSGKEAAC 494
Db 241 PEKRNPNYKESLPKCSQVQVGIKISVHKRLASEKRNPNYK-VMLASRSTSGKEAAC 300
Oy 495 EESTPSMASDHNHNAVKPEKPEPTALSPPTLS 522
Db 301 EESTPSMASDHNHNAVKPEKTAAPSPVT 328

RESULT 8
AAB43724
ID AAB43724 standard; protein; 181 AA.
XX
AC AAB43724;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated protein sequence SEQ ID NO:1169.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarteritic; antiviral;
KW antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nocitropic;
KW vasotropic; antiporiatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
XX Homo sapiens.
XX
XX WO200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US005882.
XX
XX 12-MAR-1999; 99US-0124270P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587533/55.
XX
XX N-PSDB; AAC77933.
XX

```


PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer.

XX Claim 11, Page 1790-1791; 2352pp; English.

CC AACT78607 to AACT78448 encode the human cancer associated proteins given in
 CC AAB4338 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnerable; immunomodulator;
 CC anti-diabetic; antineoplastic; antineurotic; antineurotic;
 CC anti-inflammatory; antitumor; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
 CC neurotropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies, and
 CC agonists and antagonists may be also be used in drug screens. AACT78449 to
 CC AACT8457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention

XX Sequence 181 AA:

Query Match 22.9%; Score 740; DB 3; Length 181;
 Best Local Similarity 85.1%; Pred. No. 1.6e-43;

Matches 131; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

252 GKPRCEGVNDNALYCTIGROPHNNRPMCCRCBEMFPGDVGTSBARGLLERGGDY 311

11 GKPRCEGVNDNALYCTIGROPHNNRPMCCRCBEMFPGDVGTSBARGLLERGGDY 70

312 ICPNCTIIQVODETHSATNBODSGCRSVGADGTCTSIGTVEOKSGEDQIKRIKKA 371

71 ICPNCTIIQVODETHSATNBODSGCRSVGADGTCTSIGTVEOKSGEDQIKRIKKA 130

372 NPSGKKLKIPOPVVEAFAPAPKICIGPGCSVAOP 405

131 NPSGKKLKIPOPVVEAFAPAPKICIGPGCSVAOP 164

RESULT 9

ABG22387
 ID ABG22387 standard; protein; 167 AA.

XX AC ABG22387;

DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #22378.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PP 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HISE-) HISEQ INC.

PA Drmanac RT, Liu C, Tang YT;
 PI

XX WIPI; 2001-639362/73.
 DR N-PSDB; AAS66574.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostic, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 52746; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probe, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostic as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 167 AA:

Query Match 22.7%; Score 732; DB 4; Length 167;
 Best Local Similarity 83.1%; Pred. No. 5.2e-43;

Matches 138; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

301 GRLLRNGEDYICPNCTIIQVODETHSATNBODSGCRSVGADGTCTSIGTVEOKSGED 360

1 GRLLRNGEDYICPNCTIIQVODETHSATNBODSGCRSVGADGTCTSIGTVEOKSGED 60

361 OGKRIEKAAPNSGKKLKIPOPVVEAFAPAPKICIGPGCSVAOPPSVCSNDCLKHA 420

61 OGKRIEKAAPNSGKKLKIPOPVVEAFAPAPKICIGPGCSVAOPPSVCSNDCLKHA 120

421 ATMRPLSSGKQKTPKRVKTPKPSLPKCSVQVIGIKISSVHR 466

121 ATMRPLSSGKQKTPKRVKTPKPSLPKCSVQVIGIKISSVHR 166

RESULT 10

ABP05496
 ID ABP05496 standard; protein; 108 AA.

XX AC ABP05496;

DT 24-JUN-2002 (first entry)

XX DE Human ORFX protein sequence SEQ ID NO:10974.

XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.

XX OS Homo sapiens.

Db 1009 APRRTLPVLTVAASFMRIPMAKPAKTPGALSHQOQOQLNFTRLGSPSKRISSETLCV 1068
Cc 399 -CQSVAPDQVYCNDCILHMAATMPLSSGK-----EOKTKPERK 439
Cc 1069 VCRPAPSTSSVYCEBCTIKRQAQALQAHAAATKPLPQNAQASLNNSPAKKPKKCOL 1128
Cc 440 VK-----TKPERK---FSLPKCSVOVGIKISSVHRLAKSEKRENFVKVMLASRSET 487
Cc 1129 FEDVLRADTVSKYERINVPKRSKGRVITGMAPSAN-QPRKWIQENPSPFVLSGTVQS 1167
Cc 488 SGKBAACESTPVSADHNTYNAV-----KPERKPEPTALSPTLSSKCTYHPKAGPQPSH 542
Cc 1188 ADABKRLIKGAPBAATSTSEPAVLGVAKKP---PEGPAQLS-----HPQNTTVQASH 1236
Cc 543 HUGGCLGSLSTRVGVLTVAASSLPARSRVQDASGPQVFLPS 586
Cc 1237 QLG-----ISSVRPLAKKDKKTTTPTVOAPT 1262

RESULT 12
ID AAM39234 standard; protein; 1728 AA.
Cc AAM39234;
Cc 22-OCT-2001 (first entry)
Cc Human polypeptide SEQ ID NO 2379.
Cc Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
Cc peripheral nervous system; neuropathy; central nervous system; CNS;
Cc Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
Cc amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
Cc chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
Cc leukaemia.
Cc Homo sapiens.
Cc WO200153312-A1.
Cc 26-JUL-2001.
Cc 26-DEC-2000; 2000WO-US034263.
Cc 23-DEC-1999; 99US-00471275.
Cc 21-JAN-2000; 2000US-00488725.
Cc 25-APR-2000; 2000US-00552317.
Cc 20-JUN-2000; 2000US-00598042.
Cc 19-JUL-2000; 2000US-00620312.
Cc 03-AUG-2000; 2000US-00653450.
Cc 14-SEP-2000; 2000US-00662191.
Cc 19-OCT-2000; 2000US-00693036.
Cc 29-NOV-2000; 2000US-00727344.
Cc (HYSB-) HYSBQ INC.
Cc Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
Cc Wang Y, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
Cc Zhou P, Goodrich R, Dzmanac RT;
Cc WPI; 2001-442253/47.
Cc N-ESDB; AAI58190.
Cc Novel nucleic acids and polypeptides, useful for treating disorders such
Cc as central nervous system injuries.
Cc Example 4; SEQ ID NO 2379; 10078pp; English.
Cc The invention relates to human nucleic acids (AAI57798-AAI61369) and the
Cc encoded polypeptides (AAM38642-AAM42213) with nootropic,
Cc immunosuppressant and cytostatic activity. The polynucleotides are useful
Cc in gene therapy. A composition containing a polypeptide or polynucleotide
Cc of the invention may be used to treat diseases of the peripheral nervous

Cc system, such as peripheral nervous injuries, peripheral neuropathy and
Cc localised neuropathies and central nervous system disease, such as
Cc Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
Cc lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
Cc utilisation of the activities such as: immune system suppression,
Cc Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
Cc and thrombolytic activity, cancer diagnosis and therapy, drug screening,
Cc assays for receptor activity, arthritis and inflammation, leukaemia and
Cc C.N.S disorders. Note: The sequence data for this patent did not form
Cc part of the printed specification
Cc SQ Sequence 1728 AA;
Cc Query Match 7.4%; Score 239.5; DB 4; Length 1728;
Cc Best Local Similarity 20.5%; Pred. No. 1.5e-07;
Cc Matches 101; Conservative 75; Mismatches 163; Indels 153; Gaps 17;
Cc 90 SLEDSSEPTSTYDVTVERASGVSSEIERS--GPVDSLSGK-----BHPASSSKA 139
Cc 155 NIQDDRNQSSSVSYLESKSVSKTKPVTHSKQNTTDAFKIVAAYEVTHSKTKNV 214
Cc 140 KGGEEEDTSDSDSGTLKELQN-----RLRRKR-KQSPV-----ERSLRSQNRRL 186
Cc 215 KSVKNTVPSQ-----QNFHREPVKRVKKQIDKEPKIQSCNSGVKVNQASHVL 265
Cc 187 KKRREDSAEFTSGVOI-----GSABQDRPLCKQBPB--ASQGPVSGSB----- 227
Cc 266 KKTLDQTL---VQIFPDLTSLSDKSHAHFGCLKBPHPAQGTGHVSHSQKCHKPOQ 321
Cc 228 -----TDIEMOLEGKATQGNTEBNPREAGKPRPCE----- 259
Cc 322 QAPAKTNSHVGBLEHGVBEHFEKEDKIKLKKPRKNUQPRORSKSPSLDEPPLPIPD 381
Cc 260 -----VYDPNALVYCICGPPHNRPMICCDRCCEWPHGDCVGISZA 299
Cc 382 NIATIRREGSDHSSSFESKTMPTPSKQCGFCKKPFHGRPMVGGCGRCDMFHGDVGLSLIS 441
Cc 300 RGRLEBNEDYICPNC-----TILQVQDETNGSATNBODSGCRSVGADTCTSIGT 352
Cc 442 QAGQMGEBDBKEVCVCKCABEDKTEILDPDLEMQATVPBPSGDKTM-----ECKERGL 496
Cc 353 VEOKSGE-----DQIGKRIEKAANPSGKKLKITOPVVEAPGAPKICPGCSSVAQPS 407
Cc 497 SKHTTNDRTKYIDPTVKKHVKILKRESGEGR-----NS 529
Cc 408 VYCSNDCILKHAATPRPLSSGKQKTEKPKVTKPEKPSLPKCSVOYGIKISSVHRL 467
Cc 530 SPCRNEIKKQJAPLRKNG---QVLPFRSSSEKSKR--IPKSTTVYTCGEKASKPG 583
Cc 468 ASERKENPVKVV 479
Cc 584 THEKQEMKKKKV 595
Cc

RESULT 13
AAM79892
ID AAM79892 standard; protein; 1766 AA.
Cc AAM79892;
Cc 06-NOV-2001 (first entry)
Cc Human protein SEQ ID NO 3538.
Cc Human; cytokine; cell proliferation; cell differentiation; gene therapy;
Cc vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
Cc tissue growth factor; immunomodulatory; cancer; leukaemia;
Cc nervous system disorder; arthritis; inflammation.
Cc Homo sapiens.
Cc OS
Cc PN
Cc WO200157190-A2.
Cc

Tue Apr 27 10:09:40 2004

us-09-787-016a-4.rag

Page 12

```

Db      841 SDCHDNEIKKQIATLRKKG---QPVLPRSSSESK--IPKESTVTCTGEKASKEG 894
               |||::|||::|
Qy      468 ASERKENPVKIV 479
               |||::|||
Db      895 THEKQEMKKCKV 906

```

Search completed: April 20, 2004, 21:18:22
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 23, 2004, 07:20:53 ; Search time 7067.26 Seconds
(without alignments)
12114.300 Million cell updates/sec

Title: US-09-787-016a-2

Sequence: 1 tccgtgtagctctggaat.....cttgcctcttgcgtg99 2867

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

BST:*
1: em_eestba:*
2: em_eesthm:*
3: em_eesthm:*
4: em_eestmu:*
5: em_eestov:*
6: em_eestpl:*
7: em_eestro:*
8: em_hic:*
9: gb_eest1:*
10: gb_eest2:*
11: gb_hic:*
12: gb_eest3:*
13: gb_eest4:*
14: gb_eest5:*
15: em_eestfun:*
16: em_eestom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rod:*
26: em_ges_phg:*
27: em_ges_vrt1:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2828.2	98.6	4619	AK042474	Mus muscu
2	2775.8	96.8	4651	BC060249	Mus muscu
3	1482.4	51.7	1614	AK014422	Mus muscu
4	1398.8	48.8	1640	BC029110	Mus muscu

5	1010.8	35.3	2752	11	BC014615	BC014615 Homo sapi
6	835.6	29.1	3385	11	AK044919	AK044919 Mus muscu
7	780.8	27.2	874	13	B0522670	B0522670 AGENCOURT
8	702.2	24.5	796	12	B1078113	B1078113 602872420
9	701	24.5	687	13	BY713603	BY713603 BY713603
10	665.4	23.2	667	14	CF169555	CF169555 B0815C09-
11	621.4	21.7	691	10	BB637982	BB637982 BB637982
12	617.4	21.5	632	13	BQ771134	BQ771134 UI-M-PTO-
13	607.4	21.2	623	14	CF535718	CF535718 UI-M-GHO-
14	594.6	20.7	645	13	BB636983	BB636983 BB636983
15	591.6	20.6	745	13	BY752619	BY752619 BY752619
16	591.4	20.6	639	10	BB664208	BB664208 BB664208
17	587.4	20.5	613	14	CB518822	CB518822 UI-M-GHO-
18	579.4	20.2	693	10	BB292563	BB292563 BB292563
19	570.6	19.9	690	14	CA327345	CA327345 UI-M-PYO-
20	516.8	18.0	1086	9	AL578222	AL578222 AL578222
21	494	17.2	925	9	AL546121	AL546121 AL546121
22	482.8	16.8	486	28	CC156300	CC156300 XB629 Bay
23	481.4	16.8	492	12	BG342505	BG342505 602374063
24	469.4	16.4	491	12	BG342512	BG342512 602374072
25	467.2	16.3	670	12	B1653462	B1653462 603300427
26	450.4	15.7	521	14	CP167818	CP167818 B0790B04-
27	448.2	15.6	476	28	AZ390149	AZ390149 1M0151B22
28	438.4	15.3	891	13	BX350413	BX350413 BX350413
29	428	14.9	653	10	AM373892	AM373892 QV0-BT053
30	427.6	14.9	653	10	AM373929	AM373929 QV3-BT053
31	427.6	14.9	1093	9	AL555810	AL555810 AL555810
32	424.6	14.8	1018	9	AL550675	AL550675 AL550675
33	420.2	14.7	1033	9	AL580669	AL580669 AL580669
34	415.2	14.5	495	9	AL549786	AL549786 ve78e11.y
35	412.6	14.4	506	14	CD541350	CD541350 B0231A09-
36	412.6	14.4	542	14	CD543468	CD543468 B0249C01-
37	402	14.0	760	10	BR298017	BR298017 601118492
38	398	13.9	773	12	BG545979	BG545979 602573278
39	392	13.7	1201	9	AL514138	AL514138 AL514138
40	389.8	13.6	850	14	CD364822	CD364822 UI-H-FT7-
41	389.2	13.6	772	10	BR299127	BR299127 601118879
42	386.8	13.5	650	10	BR298593	BR298593 601119753
43	386.4	13.5	868	10	BF306721	BF306721 601888796
44	385.8	13.5	1119	9	AL550431	AL550431 AL550431
45	385.2	13.4	498	12	BM247400	BM247400 K0801D11-

ALIGNMENTS

RESULT 1
AK042474
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
AK042474 4619 bp mRNA linear HTC 19-SEP-2003
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:AK0095A07 product:death inducer-oblierator-1, full
insert sequence.
AK042474
AK042474.1 GI:26335138
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci,P., Shibata,Y., Bayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Komoto,H., Akiyama,Y., Nishi,K., Kitahara,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,S., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
Riken integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
11076861

TITLE
4
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE
AUTHORS
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
AUTHORS
5 (bases 1 to 4619)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayashida,N., Hiramoto,K., Hirakawa,T., Hirozane,T.,
Hori,F., Imocani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanakawa,T.,
Kawachi,H., Kawai,J., Kojima,Y., Kondo,S., Komano,H., Koda,M.,
Koya,S., Kuribara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaiya,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Miyamatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Exploration Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://lancom.gsc.riken.go.jp/
Location/Qualifiers
1..4619
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="CS:EBL/6J"
/db_xref="PANTOM_DB:A630095A07"
/db_xref="MGI:2407519"
/db_xref="taxon:10090"
/clone="A630095A07"
/tissue_type="Chymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
299..2143
/note="unnamed protein product; death
inducer-obliterating-1 (MGI:MGI:1344352, GB|A0238332,
evidence: BLASTN, 99%, match=2599)

putative"
/codon_start=1
/protein_id="BAC31270.1"
/db_xref="GI:26335139"
/translation="MDDKGHLSNEBAPAKIPTSFRKFWGFRPTTIAKRGADTE
VDPSEQPOOHNLRLSRGQPKRTREYELTYRRRGKNVPSLSDSPSTSTV
TDVPAASGVSSESSIRISGPDVSLRSGREDSAEFTSGVGSABORPLCKQPEA
LONLRBRBRBPVBRSLRGSQNLRRKRBREDSAEFTSGVGSABORPLCKQPEA
SQGVSSSEETDINOLEKXATQNTENRPEZGRKRECVYDNPALCYICROPHN
RPMICDCEBPFHDDVGISBARGLLRNEDYICNCTILQYQDSTNGSATDOD
SGCRVAGDGTDCSTIGTIRKSGSDQIKRIEAPNPSGKKLIPOVVEAGAP
KCIQGVGSSVAQPDPSVYCSNDCLLHAAATRPLSGRQKTRPKVTRKSLP
KCSYVGGKISVHRLASREKRPVKKVMAISRTSGKBAACSTSPMSDHNYN
AVKRPKRPTRALSPTLISKCTYHPKAFPGPSHGLGCLGTRVGLVLIVAAS
SLPARSRQDASGPQVPLPSLMSLSGWLKSCVGLMLAIVSFRPM"

ORIGIN
Query Match 98.6%; Score 2828.2; DB 11; Length 4619;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2856; Conservative 0; Mismatches 8; Indels 3; Gaps 2;
1 TCCGTGTAGCTCTGGAATGCTGCGGATTCGCGCGCGGAGCTTTTAAAGG 60
68 TCCGTGTAGCTCTGGAATGCTGCGGATTCGCGCGCGGAGCTTTTAAAGG 127
61 CAGTCCCACTGCTGCTTGTGTGAGAGCTCGGATCTTCAATTGATGACTG 120
128 CAGTCCCACTGCTGCTTGTGTGAGAGCTCGGATCTTCAATTGATGACTG 187
121 ACTGCAACAGTGGGGTGGAGCTTGGCGGTCTGCTGACCTGGCCCAAGTCTATATTT 180
188 ACTGCAACAGTGGGGTGGAGCTTGGCGGTCTGCTGACCTGGCCCAAGTCTATATTT 247
181 TATGTAGATTTCCAGCCAAAGTTTCCAGCTTTCAAGTTTGGACAGTATGAT 240
248 TATGTAGATTTCCAGCCAAAGTTTCCAGCTTTCAAGTTTGGACAGTATGAT 307
241 AAGGGCACTGAGCAATGAGAGACCAAGGCTATCAACCCACAGTAAGAGTTT 300
308 AAGGGCACTGAGCAATGAGAGACCAAGGCTATCAACCCACAGTAAGAGTTT 367
301 AAGAAACCTGGGGTTTTCAGAAACCAAGATTCGCAACGTAAGGTCACAGACAG 360
368 AAGAAACCTGGGGTTTTCAGAAACCAAGATTCGCAACGTAAGGTCACAGACAG 427
361 GAGGCGAATCCGAGTGAAGACCAACGACGATTAACCTTCTGCGCGCGAGTGA 420
428 GAGGCGAATCCGAGTGAAGACCAACGACGATTAACCTTCTGCGCGCGAGTGA 487
421 CGGCAACCAAAAGTACGAGAGGTTAGAAAGTTTCTTACACAGGTTGCGCGAGG 480
488 CGGCAACCAAAAGTACGAGAGGTTAGAAAGTTTCTTACACAGGTTGCGCGAGG 547
481 AAAAAAATGTCGCGGTGTCCTGAGAGATTCAGTAGACCCACATCTTCAAGTCACT 540
548 AAAAAAATGTCGCGGTGTCCTGAGAGATTCAGTAGACCCACATCTTCAAGTCACT 607
541 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGCTTGAAGATTCGAAGTGGCCCT 600
608 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGCTTGAAGATTCGAAGTGGCCCT 667
601 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGTGA 660
668 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGTGA 727
661 GAGGAAGAAGACCTCTGACAGTGAAGTGAAGCTTGAAGAGTGAAGTGAAGTGA 720
728 GAGGAAGAAGACCTCTGACAGTGAAGTGAAGCTTGAAGAGTGAAGTGAAGTGA 787
721 CGCTTCGAGAAAGGAGAGCAAAAGCTTGGAGAGGCTTCTGAGAGGAGTGAAGT 780
788 CGCTTCGAGAAAGGAGAGCAAAAGCTTGGAGAGGCTTCTGAGAGGAGTGAAGT 847

ACCESSION BC060249
 VERSION BC060249.1 GI:37805411
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM
 SOURCE Mus musculus
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 4651)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Schein, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bat, N.K., Hopkins, R.P., Jordan, H., Moore, T., Max, S.I., Wang, D., Heish, P., Diachenko, L., Marsina, K., Farmer, A., Rubin, G.M., Hong, L., Staelen, T.M., Soares, M.B., Donald, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carinici, P., Prange, C., Kane, S.S., Loggellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huijib, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, D.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.B., Scherch, A., Schein, J.B., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 4651)
 Strausberg, R.
 Direct Submission
 Submitted (20-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: http://mgc.ncl.nih.gov
 Contact: MGC help desk
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 CDNA Library Preparation: M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ram Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutische, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Matheson, Candice McLeary, Steven Ness, Pawan Pandon, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smal, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranda Teal, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
 Series: IRAC Plate: 126 Row: k Column: 20
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency OR analysis. Similarly but not identity to protein
 This clone has the following problem: frame shifted.
 Location/Qualifiers
 1. 4651
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6826043"
 /tissue_type="Brain, mouse, 13.5, 14.5, 16.5, 17.5 dpc"
 /clone_id="NIH_BMAP_FY0"
 /lab_host="DH10B"

ORIGIN /note="Vector: pYX-Asc"
 Query Match 96.8%; Score 2775.8; DB 11; Length 4651;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 2814; Conservative 0; Mismatches 7; Indels 4; Gaps 3;
 44 GGAGCTGTTTAAAGGAGGAGTCCCACTGCTTGGTGTGGAAGTCCGGAATCTT 103
 90 GAGAGCTGTTTAAAGGAGGAGTCCCACTGCTTGGTGTGGAAGTCCGGAATCTT 149
 104 CTCATTGATGAACCTGACTGACCAACAGTGGGTGAGGCTTGGCTTGGTACCTGG 163
 150 CTCATTGATGAACCTGACTGACCAACAGTGGGTGAGGCTTGGCTTGGTACCTGG 209
 164 CCCAGGCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 223
 210 CCCAGGCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 269
 224 GGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 283
 270 GGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 329
 284 CCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 343
 330 CCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 389
 344 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 403
 390 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 449
 404 CCCTGCGCGCAGTGAACCGGCAACCAAAAGTAAAGGATTTCTTAACA 463
 450 CCCTGCGCGCAGTGAACCGGCAACCAAAAGTAAAGGATTTCTTAACA 509
 464 CGGTTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 523
 510 CGGTTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 569
 524 CATCTTCCACAGTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 583
 570 CATCTTCCACAGTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 629
 584 AGATCAGAAAGTGGCCCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 643
 630 AGATCAGAAAGTGGCCCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 689
 644 AGGCAAAAGAGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 703
 690 AGGCAAAAGAGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 749
 704 TGAAGGAACCTTCAAGACCCCTTGGAGGAAGGAGGAGGAGGAGGAGGAGGAGG 763
 750 TGAAGGAACCTTCAAGACCCCTTGGAGGAAGGAGGAGGAGGAGGAGGAGGAGG 809
 764 TGAAGGAACCTTCAAGACCCCTTGGAGGAAGGAGGAGGAGGAGGAGGAGGAGG 823
 810 TGAAGGAACCTTCAAGACCCCTTGGAGGAAGGAGGAGGAGGAGGAGGAGGAGG 869
 824 GGAAGTCCCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 883
 870 GGAAGTCCCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 929
 884 CTAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 943
 930 CTAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 989
 944 AGGCACTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1003
 990 AGGCACTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1049
 1004 GTGAGGTTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1063

Db 1050 GTGAGGTTTAAAGACCCCAATGCGCTGTATCTGATCTGCGCGACGCTCAACAACAGGT 1109
Qy 1064 TTATGATCTGCTGATCGATGATGAGAGAGTGTTCATATGATGATGATGATGATTTCTG 1123
Db 1110 TTATGATCTGCTGATCGATGATGAGAGTGTTCATATGATGATGATGATGATTTCTG 1169
Qy 1124 AGGCGGAGAGGCGGCTCTGTGAAAGAAAGGAGAAAGACTATCATCTGCCAATTTGACCA 1183
Db 1170 AGGCGGAGAGGCGGCTCTGTGAAAGAAAGGAGAAAGACTATCATCTGCCAATTTGACCA 1229
Qy 1184 TTTTTCAGATGACAGATGAGCAAAAGGTTAGCGCCCAATAGAGAGAGACTCTGGTGCA 1243
Db 1230 TTTTTCAGATGACAGATGAGCAAAAGGTTAGCGCCCAATAGAGAGAGACTCTGGTGCA 1289
Qy 1244 GATCTGAGGATGATGATGAGCAAGCTGCAAGAGATAGGAGCAATGAGAGAGAGTCCG 1303
Db 1290 GATCTGAGGATGATGATGAGCAAGCTGCAAGAGATAGGAGCAATGAGAGAGAGTCCG 1349
Qy 1304 GAGAGAGCAAGGAGCAATAAAGGATGAGATTTGAGAAAGGAGCAAAAGGCGGCAAGAA 1363
Db 1350 GAGAGAGCAAGGAGCAATAAAGGATGAGATTTGAGAAAGGAGCAAAAGGCGGCAAGAA 1409
Qy 1364 AACTCAAGATTTTCAAGCTCTCTGTAAGAGGCTCTCTGATGATGATGATGATGATGAT 1423
Db 1410 AACTCAAGATTTTCAAGCTCTCTGTAAGAGGCTCTCTGATGATGATGATGATGATGAT 1469
Qy 1424 GGTGTTCAAGTATGAGCAAGCTGATCTGTGATTTGAGATTAATGATGATTTCAAAAC 1483
Db 1470 GGTGTTCAAGTATGAGCAAGCTGATCTGTGATTTGAGATTAATGATGATTTCAAAAC 1529
Qy 1484 ACGGAGAGCTACCAATGAGATTTCTAAGTTCAAGTTAAGAAAC-AAAAAACAACCCAG 1542
Db 1530 ACGGAGAGCTACCAATGAGATTTCTAAGTTCAAGTTAAGAAACAAAAAACAACCCAG 1589
Qy 1543 GAAAAAGTCAAGAGAGAGAGGAAAAAGTTCAAGTTTCCAAAAATGAGATTTCAAGTTGAG 1602
Db 1590 GAAAAAGTCAAGAGAGAGAGGAAAAAGTTCAAGTTTCCAAAAATGAGATTTCAAGTTGAG 1649
Qy 1603 ATTAAATCTCTCTGATGAGCAAGAGCTAGCGGTCAAGAGAAAGGAAAAAGGCAATGAG 1662
Db 1650 ATTAAATCTCTCTGATGAGCAAGAGCTAGCGGTCAAGAGAAAGGAAAAAGGCAATGAG 1709
Qy 1663 AAGATGATGAGCTTCCAGAGATGAGACTTCTGAGAAAGAGGAGGCTGTGAGAGCAAC 1722
Db 1710 AAGATGATGAGCTTCCAGAGATGAGACTTCTGAGAAAGAGGAGGCTGTGAGAGCAAC 1769
Qy 1723 ACACATCTCTGAGCAAGTGAACAACAATGATGATGAGAGGAGAGAGGAGAGAGAG 1782
Db 1770 ACACATCTCTGAGCAAGTGAACAACAATGATGATGAGAGAGGAGAGAGAGAGAG 1829
Qy 1783 CCACTGCACTCTGAGCAAGGCTATTTAGTAAATGATGATGATGATGATGATGATGATGAT 1842
Db 1830 CCACTGCACTCTGAGCAAGGCTATTTAGTAAATGATGATGATGATGATGATGATGATGAT 1889
Qy 1843 CCAAGGCGGCTCCCATCATCTGAGGAGGCTGAGGCTGTGATGAGCAAGAGTCTGAGGT 1902
Db 1890 CCAAGGCGGCTCCCATCATCTGAGGAGGCTGAGGCTGTGATGAGCAAGAGTCTGAGGT 1949
Qy 1903 GTTCTGATGATGATGAGCAAGGCTGATGATGATGATGATGATGATGATGATGATGAT 1962
Db 1950 GTTCTGATGATGATGAGCAAGGCTGATGATGATGATGATGATGATGATGATGATGAT 2009
Qy 1963 TCTGAGACCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2022
Db 2010 TCTGAGACCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2069
Qy 2023 TGTGAGGCTCTCATGTTGAGAGCAATTTCTAATTTCAAGTTTGAAGCTTGTGATGATTT 2082
Db 2070 TGTGAGGCTCTCATGTTGAGAGCAATTTCTAATTTCAAGTTTGAAGCTTGTGATGATTT 2129
Qy 2083 GAACTGCTCTCAACAAGATGAGCAACAATCAATGATGATGATGATGATGATGATGATGAT 2142
Db 2130 GAACTGCTCTCAACAAGATGAGCAACAATCAATGATGATGATGATGATGATGATGATGAT 2189

Qy 2143 AGTTCTCAACAAGAGGCTCCAGCAACAACATGAGATTTTCCAGATTAACCTATTTTCT 2202
Db 2190 AGTTCTCAACAAGAGGCTCCAGCAACAACAT- GAGTTTCCAGATTAACCTATTTT--C 2246
Qy 2203 TCAACATGATGAGGCTCTCTGATGAGAAAGATTTTCTATGATGATGATGATGATGATGAT 2262
Db 2247 TCAACATGATGAGGCTCTCTGATGAGAAAGATTTTCTATGAGAAAGATTTTCTATGATGAT 2306
Qy 2263 TTTGAAAGAGAGGATGAGCAATTTGATGATTTTCAATTTCACTACTACACACTACATAAACG 2322
Db 2307 TTTGAAAGAGAGGATGAGCAATTTGATGATTTTCAATTTCACTACTACACACTACATAAACG 2366
Qy 2323 GTTGAATTTGAAGATCATCTGCTGCTCAAGATCCCATGATGATGATGATGATGATGATGAT 2382
Db 2367 GTTGAATTTGAAGATCATCTGCTGCTCAAGATCCCATGATGATGATGATGATGATGATGAT 2426
Qy 2383 ACACCTACAGAGATTTCTCAATGAGAGATTTGAGAGATTTGAGAGATTTGAGAGATTTGAG 2442
Db 2427 ACACCTACAGAGATTTCTCAATGAGAGATTTGAGAGATTTGAGAGATTTGAGAGATTTGAG 2486
Qy 2443 GATGCGGAGATTTCAATCAAGTGTGATTTAGTTGATGATGATTTGATGATGATTTGATGAT 2502
Db 2487 GATGCGGAGATTTCAATCAAGTGTGATTTAGTTGATGATGATTTGATGATGATTTGATGAT 2546
Qy 2503 GATCACTGAGATTTGAGGCAATTTCTGATTTCACTGCTGATCAAGATGACCTTTCTACTT 2562
Db 2547 GATCACTGAGATTTGAGGCAATTTCTGATTTCACTGCTGATCAAGATGACCTTTCTACTT 2606
Qy 2563 TTTGACAGATGATTTCAAGTAAATTTTCACTGATTTTCAAGATGATGATTTTCAAGTAA 2622
Db 2607 TTTGACAGATGATTTCAAGTAAATTTTCACTGATTTTCAAGATGATGATTTTCAAGTAA 2666
Qy 2623 CTCTGACAGGCTCAAGGCTCAAGTAAAGGAGGCTCATGAGGCTCAAGGCTCAAGGCTCAAG 2682
Db 2667 CTCTGACAGGCTCAAGGCTCAAGTAAAGGAGGCTCATGAGGCTCAAGGCTCAAGGCTCAAG 2726
Qy 2683 CTGTAAGTTGAGGAGGCTCTCTGCGCCCACTGTAAGATGATGATGATGATGATGATGAT 2742
Db 2727 CTGTAAGTTGAGGAGGCTCTCTGCGCCCACTGTAAGATGATGATGATGATGATGATGAT 2786
Qy 2743 CACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2802
Db 2787 CACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2846
Qy 2803 GCTGAGATTTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2862
Db 2847 GCTGAGATTTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2906
Qy 2863 CTGAG 2867
Db 2907 CTGAG 2911

RESULT 3
AK014422 1614 bp mRNA linear HTC 20-SEP-2003
LOCUS AK014422
DEFINITION Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830408B01 product:death inducer-obliterator-1, full insert sequence.
ACCESSION AK014422
VERSION AK014422.1 GI:12852263
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253

PUBMED REFERENCE	10349636
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBMED	Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasakii,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakauchi,S., Ikegami,T., Kasaiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T. and Matsubara,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kire,A. and Hayashizaki,Y.
JOURNAL MEDLINE PUBMED	Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED	Nature 409, 685-690 (2001) 5
REFERENCE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 1614)
TITLE	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kaubekawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuana,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ono,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shidara,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyai,T., Yamamura,T., Yasunishi,Y., Yoshida,K., Yoshino,M., Yuramatsu,M. and Hayashizaki,Y.
JOURNAL MEDLINE PUBMED	Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel.81-45-503-9222, Fax:81-45-503-9216]
REFERENCE	COMMENTS
TITLE	COMMENT
JOURNAL	Further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGACAGATTCCTCAGTATTAAATTAATCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence 15'. GAGAGACAGATTCCTCAGTATTAAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified plasmidscript KS(+) after bulk excision from lambda PhiC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES	Location/Qualifiers
SOURCE	1..1614

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:3830408B01"
/db_xref="MGI:1910351"
/db_xref="taxon:10090"
/clone="3830408B01"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18 days pregnant adult"
31.. >1614
/node="death inducer-obliterator-1 (MGD|MGI:1344352,
GB|AJ238332, evidence: BLASTN, 95%, match=2599)
putative"

ORIGIN

Query Match
Best Local Similarity 99.4% Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 6; Indels 3; Gaps 2;
51.7%; Score 1482.4; DB 11; Length 1614;

```

Query Match	51.7%	Score 1482.4	DB 11	Length 1614	
Best Local Similarity	99.4%	Pred. No. 0			
Matches 1509	Conservative	0	Mismatches 6	Indels 3	Gaps 2
QY	1	TCCTGTAGACTCTGGAATGCTGCGGATCCCGCGGCGGGAGCTTTTAAGAGG	60		
DB	100	TCCGTGTAGACTCTGGAATGCTGCGGATCCCGCGGCGGGAGCTTTTAAGAGG	159		
QY	61	CAGTCCCACTGTCCTTTGTTGTTCCAAAGCTCCGGAACTCTCATTTGATGACCTG	120		
DB	160	CAGTCCCACTGTCCTTTGTTGTTCCAAAGCTCCGGAACTCTCATTTGATGACCTG	219		
QY	121	ACTGCACAAGTGGGGTGAAGGCTTGAGCGCTTGCTTGAACCTGAGCCCAAGTCTATATTT	180		
DB	220	ACTGCACAAGTGGGGTGAAGGCTTGAGCGCTTGCTTGAACCTGAGCCCAAGTCTATATTT	279		
QY	181	TATGTAGATTTCCAGCCAAAGGTTTCCAAAGCTTTCAAGTTTGGACAGGTAATGAT	240		
DB	280	TATGTAGATTTCCAGCCAAAGGTTTCCAAAGCTTTCAAGTTTGGACAGGTAATGAT	339		
QY	241	AAAGGGCACTTGACGATGAGAACCAATCCAGGCTATCAAAACCCACAGTAAAGAGTTC	300		
DB	340	AAAGGGCACTTGACGATGAGAACCAATCCAGGCTATCAAAACCCACAGTAAAGAGTTC	399		
QY	301	AGGAAAACCTGGGGGTTTTCGAAAGACCAAGATTGCGAAACGTGAGGGTCAAGAGACAG	360		
DB	400	AGGAAAACCTGGGGGTTTTCGAAAGACCAAGATTGCGAAACGTGAGGGTCAAGAGACAG	459		
QY	361	GAGCGGACCCCACTGAGCAGCAACCAACGACGATTAACCTTCCCTGCGCGCACTGGA	420		
DB	460	GAGGTGACCCCACTGAGCAGCAACCAACGACGATTAACCTTCCCTGCGCGCACTGGA	519		
QY	421	CGGCAACCAAAACGTACGTAGAGGGTTGAAGGTTTCTTAACACAGGTTGCGGCGCGAGGG	480		
DB	520	CGGCAACCAAAACGTACGTAGAGGGTTGAAGGTTTCTTAACACAGGTTGCGGCGCGAGGG	579		
QY	481	AAAAAGATGTGCGCGGTCCCTGAGAGATTCCAGTGAAGCCACATCTTTCACAGTCACT	540		
DB	580	AAAAAGATGTGCGCGGTCCCTGAGAGATTCCAGTGAAGCCACATCTTTCACAGTCACT	639		
QY	541	GATGTGAGACACGCTTCCGAGGGGACGTTGAACAAGTTTCAGATTCAGAGTGGGCCCT	600		
DB	640	GATGTGAGACACGCTTCCGAGGGGACGTTGAACAAGTTTCAGATTCAGAGTGGGCCCT	699		
QY	601	GTAATGACTCCTTTGAGGAAAGAACATCTGCTCTTTCGAAAAGCCAAAGAGGTGA	660		
DB	699	GTAATGACTCCTTTGAGGAAAGAACATCTGCTCTTTCGAAAAGCCAAAGAGGTGA	756		
QY	661	GAGGAAGAAACACTCTGACAGTGAACAGTGAATGACCTTAAGTTGAAGAACTTCAAGAAC	720		
DB	757	GAGGAAGAAACACTCTGACAGTGAACAGTGAATGACCTTAAGTTGAAGAACTTCAAGAAC	816		
QY	721	CGCCTTCGAGAAACCGAGAGCAACACTGTGAGAGGCTCCCTGAGAGGCACTGAAAT	780		
DB	817	CGCCTTCGAGAAACCGAGAGCAACACTGTGAGAGGCTCCCTGAGAGGCACTGAAAT	876		


```

Db      121 GGGGTGAGGCTTGGCCGCTGCTGAGACCTGAGCCCGAGGCTATAATATTTATGAGATTT 180
Qy      193 CCAGCCAAAGTTTCCAGCTTTGAGTTTGGAC----- 227
Db      181 CCAGCCAAAGTTTCCAGCTTTGAGTTTGGACAGGGGTTTTGTTTTATGCGAGTCT 240
Qy      228 ----- 227
Db      241 CCCCCTAAGATGGAACATGACGATTAACCTTTCATTACTAGCATTTAATGCAAGTTAAAG 300
Qy      228 -----AGTATGATGATTAAGGGCACTTGAGCAATGA 260
Db      301 CCTGAGACGAGGAAAAGATTTCTCAGGATATGATTAAGGGCACTTGAGCAATGA 360
Qy      261 GGAAGCACTCCAGGCTATCAAAACCCAGATTAAGGATTCAGGAAACCTGGGGTTTTG 320
Db      361 GGAAGCACTCCAGGCTATCAAAACCCAGATTAAGGATTCAGGAAACCTGGGGTTTTG 420
Qy      321 AAGAACAAGATTTGCAAAAGGAGTGCAGAGAGACAGAGAGCGGAGCCCAAGTGA 380
Db      421 AAGAACAAGATTTGCAAAAGGAGTGCAGAGAGACAGAGAGCGGAGCCCAAGTGA 480
Qy      381 GCAACCAAGAGAGATTAACCTTCCCTGCGCGAGTGAAGCGCAACCAAAAGTACTGA 440
Db      481 GCAACCAAGAGAGATTAACCTTCCCTGCGCGAGTGAAGCGCAACCAAAAGTACTGA 540
Qy      441 GAGGTTGAGAGATTTCTTACACAGGTTCCGCGCGAGGAGGAAAAGATTTGCGGTTTC 500
Db      541 GAGGTTGAGAGATTTCTTACACAGGTTCCGCGCGAGGAGGAAAAGATTTGCGGTTTC 600
Qy      501 CCTGAGAGATTCAGATGAGCCCACTCTTCCACAGTCACTGATGTGAGAGACGTTCCGA 560
Db      601 CCTGAGAGATTCAGATGAGCCCACTCTTCCACAGTCACTGATGTGAGAGACGTTCCGA 660
Qy      561 GGGAGGCTTTGAAAGCAGTTCTGAGATCAAGATGAGGCGCTGATCTGACTTTAGGAA 620
Db      661 GGGAGGCTTTGAAAGCAGTTCTGAGATCAAGATGAGGCGCTGATCTGACTTTAGGAA 720
Qy      621 AGAATCTCTGCTCTTCTTCTGAAAAGGCAAAAGAGGTGAAGGAGAAAGACCTCTGA 680
Db      721 AGAATCTCTGCTCTTCTTCTGAAAAGGCAAAAGAGGTGAAGGAGAAAGACCTCTGA 780
Qy      681 CAGTGAACGATGAGCCTTTACGTTGAAAGAACTTCAGAACCGCTTCCGAGAGAACGAGA 740
Db      781 CAGTGAACGATGAGCCTTTACGTTGAAAGAACTTCAGAACCGCTTCCGAGAGAACGAGA 840
Qy      741 GCAAGAACCTGTGAGAGAGTCCCTGAGAGGAGTCAAGATTCCTGAGAGAGAGCGCAG 800
Db      841 GCAAGAACCTGTGAGAGAGTCCCTGAGAGGAGTCAAGATTCCTGAGAGAGAGCGCAG 900
Qy      801 AGAGGAATTTCTGCGGAAAACGAGAGTGTCCAAATAGGAGTGCAGAGCAGAGACAGACC 860
Db      901 AGAGGAATTTCTGCGGAAAACGAGAGTGTCCAAATAGGAGTGCAGAGCAGAGACAGACC 960
Qy      861 TCTCTGTAAGCAGAGCCTGAGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAGATGA 920
Db      961 TCTCTGTAAGCAGAGCCTGAGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAGATGA 1020
Qy      921 CATAGAAAATCAGTTGAAAGGAGAGCGCATCAGGAGAAATACAGAGAGAAACCCCAAGGA 980
Db      1021 CATAGAAAATCAGTTGAAAGGAGAGCGCATCAGGAGAAATACAGAGAGAAACCCCAAGGA 1080
Qy      981 AGCGGGCAAAACCAAGCCTGAGATGAGAGTTTAAAGACCCCAATGCTGTACTGATCTG 1040
Db      1081 AGCGGGCAAAACCAAGCCTGAGATGAGAGTTTAAAGACCCCAATGCTGTACTGATCTG 1140
Qy      1041 CCGCAGGCTCAACAACAAGGTTATATCTGCTGTGATCGGTGTGAGAGTGTGTTCA 1100
Db      1141 CCGCAGGCTCAACAACAAGGTTATATCTGCTGTGATCGGTGTGAGAGTGTGTTCA 1200
Qy      1101 TGGTGAATGTGTGGTATTTCTGAGAGCCCGAGAGGCGCTCTCTGAGAAAGAAAGGGAGAA 1160
Db      1201 TGGTGAATGTGTGGTATTTCTGAGAGCCCGAGAGGCGCTCTCTGAGAAAGAAAGGGAGAA 1260

```

```

Qy      1161 CTACATCTGCCCAATATTCACCATTTTTCAGATGAGAGATGAGACAAACGATAGCCAC 1220
Db      1261 CTACATCTGCCCAATATTCACCATTTTTCAGATGAGAGATGAGACAAACGATAGCCAC 1320
Qy      1221 CAATAGCAGAGACTCTGGGTGAGATCTGTGGGTCTCTGATGCAACAGACTGCAAGACT 1280
Db      1321 CAATAGCAGAGACTCTGGGTGAGATCTGTGGGTCTCTGATGCAACAGACTGCAAGACT 1380
Qy      1281 AGGAGCAGTAGAGAGAGTCCGAGAGAGACAGAGGCAATTAAGGATGAGTAAGGC 1340
Db      1381 AGGAGCAGTAGAGAGAGTCCGAGAGAGACAGAGGCAATTAAGGATGAGTAAGGC 1440
Qy      1341 AGCAAAACCCAGCGGCAAGAAAACCTCAAGATATTTCCAGCTGTGATGAGGCTCTGG 1400
Db      1441 AGCAAAACCCAGCGGCAAGAAAACCTCAAGATATTTCCAGCTGTGATGAGGCTCTGG 1500
Qy      1401 TGCTCTTAATGATGATGCGCTGAGTGTTCACATGATGACAGCTGATCTGTGATATG 1460
Db      1501 TGCTCTTAATGATGATGCGCTGAGTGTTCACATGATGACAGCTGATCTGTGATATG 1560
Qy      1461 CAGTATGATGATGATGCTCAAAACGAGAGAGCTACATGAGATTTCTAGTTAGGTTAA 1520
Db      1561 CAGTATGATGATGATGCTCAAAACGAGAGAGCTACATGAGATTTCTAGTTAGGTTAA 1620
Qy      1521 AGAACAACCAAAACAAA 1536
Db      1621 AGAACAACCAAAACAAA 1636

RESULT 5
BC014615
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCES
AUTHORS
JOURNAL
REMARK

```

BC014615 2752 bp mRNA linear HTC 17-DEC-2003

Homo sapiens cDNA clone IMAGE:3532298, containing frame-shift errors.

BC014615

BC014615.2 GI:33869935

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 2752)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.P., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, L., Diatchenko, L., Marziani, K., Farmer, A.A., Rubin, G.M., Hong, L., Stappleton, M., Soares, M.B., Bonaldo, M.F., Cabavanc, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshilyuk, S., Carrinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Bulay, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Hellon, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmitz, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.B., Butterfield, A., Schein, J.B., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2752)

Strausberg, R.

Direct Submission

Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT On Aug 19, 2003 this sequence version replaced gi:15779097.

Contact: MGC help desk
Email: c9ab8-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbio.org
Anup Madan, Jessica Pahey, Brin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 11 Row: 3 Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18375616
This clone has the following problem: frame shifted.

FEATURES
source Location/Qualifiers

1..2752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGB:353228"
/cissue_type="Muscle, rhabdomyosarcoma"
/clone_id="NH108-MGC 17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Query Match 35.3%; Score 1010.8; DB 11; Length 2752;
Best Local Similarity 71.8%; Pred. No. 3.2e-236;
Matches 1481; Conservative 0; Mismatches 532; Indels 51; Gaps 10;
16 GAATGGCTGGCGGATCCCGGGCGCGGAGACTTTTAAGAGGAGTCCCACTGTC 75
1 GAAATGGCTGGCGGAGCTTGAAGAGGCTTGAAGAGGCTTGAAGAGGCTTGAAG 60
76 CTTGGTGTTCGAAAGCTCCGATCTTCATTTGATGATGATGATGATGATGATGAT 135
61 AATCGAGTGTGAAATGAAATGAGAGCTTCTGATGATGATGATGATGATGATGAT 120
136 GT-GAGGCTTGGCGCTGCTTG-----ACCTGCCCGCAAGCTTAATATTT 181
121 GTTGAAGCTGCTGCTGCTGCTGACCCAGAGTTCGCTGACGAGGCTTTTGCTGT 180
182 ATGTAGATTTCCAGCCAAAGTTTCCAGCTTTCAGTTTGGAGAGAGTATGATGAT 241
181 ATTAGATTT-CAGGAAAGGTTCAGCTTTCAGTGTGAGAGAGTATGATGATGAT 239
242 AAGGCACTGAGCAATGAGAGAGCAAGCAAGCTTCAAAACCCAGTAAAGAGTTCA 301
240 AAGGCACTGAGCAATGAGAGAGCAAGCAAGCTTCAAAACCCAGTAAAGAGTTCA 299
302 GGAAGAACTGGGGTTTGAAGAAACAGATTTGCCAAAGTGAAGGTGAGAGAGAG 361
300 GGAAGAACTGGGGTTTGAAGAAACAGATTTGCCAAAGTGAAGGTGAGAGAGAG 359
382 AGGCGGAGCCCGAGTGAAGCAAGC-----ACAGAGATTAACCTTCCCTGGCGGCA 415
360 AGGCGGAGCCCGAGTGAAGCAAGC-----ACAGAGATTAACCTTCCCTGGCGGCA 419
416 GTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 475
420 GTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 479
476 GAGGAGAAAGAAATGTGCGGTGTCTTGAAGATTCAGATGAGCCCAATCTTCAAG 535
480 GGGGAGGAGGAGAGCTGTCTCTCTGAGAGATTTCTGTGAGCCCAAGTCTGCTCG 539
536 TCACGATGTGAGAGAGCTTCCAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 595
540 CCACAGAGCCCGAGAGAGCTCCGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAG 599

QY 596 GCCCTGATCTGATCTCTTA--GGAGAGACATCTGCTCTTCTGAAAGGCAAG 652
DB 600 GCCCCAGTCTGCTTCACAGCTGTGAAAGAGACAGCTCTCTGAAAGGTAAG 659
QY 653 GAGGTGAAGAGAGAGAGACCTTGAAGAGTGAAGAGTGAAGAGTGAAGAGAG 712
DB 660 GAGGTGAAGAGAGAGAGACCTTGAAGAGTGAAGAGTGAAGAGTGAAGAGAG 719
QY 713 TTCAAGACCGCTTCGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 772
DB 720 TTCAAGATGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
QY 773 GTCAAGATGCTTCGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 832
DB 780 TCAGAGTGGCTTCGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
QY 833 AATTAGCAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 892
DB 840 AGGCAAGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 899
QY 893 GACAGATGCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 952
DB 900 GGGTGTGCTCCAGAGCTGGAGAAAGATGACAGAGAGAGTATGAGAGAGAGAG 959
QY 953 AGGAAATACAG 1012
DB 960 AGGACATCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1019
QY 1013 AGAGCCCAAGTCCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1072
DB 1020 AGAGCCCAAGTCCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1079
QY 1073 GCTGTATCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1132
DB 1080 GCTGTATCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1139
QY 1133 GCGGCTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1192
DB 1140 GAGGCTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1199
QY 1193 TGAAGATGAG 1252
DB 1200 TGAAGATGAG 1259
QY 1253 GTGATGATGAG 1312
DB 1260 ATGCTATGAG 1319
QY 1313 AGGCAATTAAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1372
DB 1320 AAGGATTAAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1379
QY 1373 TATTCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1432
DB 1380 TCTTCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1439
QY 1433 GTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1492
DB 1440 AGTGGCGAGCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1499
QY 1493 CTACATGAGATTTCTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1552
DB 1500 CCAGATTAAGATTTCTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1559
QY 1553 AG 1612
DB 1560 AGATGAAGCAAG 1619
QY 1613 CTTGTGCAAG 1668
DB 1620 CTTGTGCAAG 1679

QY 1669 ATGCTGCTTCAGAGTGAAGACTTCTGGAGAGGAGGAGCCTGTGAGAGCAGACACCA 1728
 DB 1680 GTGCTCCCTGCGCGAGTGAAGCACTCGGAGAGGAGAGGAGCTTGTGAGAGCAGCAGC 1739
 QY 1729 TCTGAGGAGTGAAGCAAGTGAAGTGAAGGAGAGGAGAGGAGAGGAGAGGAGAGG 1788
 DB 1740 TCGTGGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1790
 QY 1789 GCACTCTGCGCCAGCCTTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1848
 DB 1791 GCTCCCTGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1850
 QY 1849 CCTCCCTGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1898
 DB 1851 CCTCCCTGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1910
 QY 1899 GAGTCTGCTGCTGATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1954
 DB 1911 TTGTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1970
 QY 1955 AAGATGCTCTGAGACCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2014
 DB 1971 AGGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030
 QY 2015 TAAAGAGCTGTAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2038
 DB 2031 TCCAAAGCTGTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2054

RESULT 6
 AK044919 3385 bp mRNA linear HTC 20-SEP-2003
 DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
 induced library, clone:B13001M08 product:death
 inducer-oblierator-1, full insert sequence.
 AK044919
 VERSION AK044919.1 GI:26336914
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsumi, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 JOURNAL sequencing pipeline with 384 multichannel sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 PUBMED 20530913
 REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)
 REFERENCE
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 JOURNAL of 60,770 full-length cDNAs
 REFERENCE Nature 420, 563-573 (2002)
 AUTHORS 6 (bases 1 to 3385)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furumasa, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kaga, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kondo, H., Kondo, M.,
 Koyama, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohse, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanita, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshinori Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-resseq@riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. Tomohiro Kono (Department of Animal
 Science, Tokyo University of Agriculture, 1377 Hnako Atsugi City,
 Kanagawa Prefecture, Japan) whose assistance we gratefully
 acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/

FEATURES
 source
 1..3385
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:B13001M08"
 /db_xref="MGI:2410273"
 /db_xref="taxon:10090"
 /clone="B13001M08"
 /tissue_type="parthenogenote"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="9.5 days embryo"
 2443..3276
 /note="unnamed protein product; death
 inducer-oblierator-1 (MGI:1344352, GB|A238332,
 evidence: BLASTN, 99%, match=2599)
 putative"
 /codon_start=1
 /protein_id="BAC32141.1"
 /db_xref="GI:26336915"
 /translation="MDKGLHNSBAPNAIKPTSKFRFTWGFRTTAKRBAQDTE
 VPSRQPOOHNLILRSGRQPKTERVEFLTYARRRCKKVPVSLBDSSEPTSTV
 TDPSRSGSVSSIRSGPDSIGKHPASBRKAGSEEDTSDSDGLTLKE
 LONRLRRRQGPVRSIRGSONRLPKRRRERDSVQSGAARDPLCKQBPBA
 SGGPVSQSDTDIDENQLBSKATQGTENPRAGKPRPCRYDPRALYICICRPPNN
 R"

ORIGIN
 Query Match 29.1%; Score 835.6; DB 11; Length 3385;
 Best Local Similarity 98.4%; Pred. No. 2.6e-193;
 Matches 844; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 211 GCTTTCAGTTTGGAGCAGATGATGATGAAGGAGCAGCTGAGCAATAGAGAGCAACC 2710


```

Db      2422 GATTGCTTTTGTCTATAGTATGATATGATTAAGGCGACCTGATGCAATAGAGAGACCTC 2481
Qy      271  AAGGCTATCAAAACCCACAGTAAGAGTTCAAGAAAACCTGGGGTTTTCAGAAACACAG 330
Db      2482 AAGGCTATCAAAACCCACAGTAAGAGTTCAAGAAAACCTGGGGTTTTCAGAAACACAG 2541
Qy      331  ATTGCAAAACGTGAGGTTGACAGAGACACAGAGCGGACCCAGTAGAGACAAACACAG 390
Db      2542 ATTGCAAAACGTGAGGTTGACAGAGACACAGAGCGGACCCAGTAGAGACAAACACAG 2601
Qy      391  CAGCAATACCTTCCCTGCGCGCGCAGTGAACCGCAACCAAAACCTACTGAGAGGTTGAA 450
Db      2602 CAGCAATACCTTCCCTGCGCGCGCAGTGAACCGCAACCAAAACCTACTGAGAGGTTGAA 2661
Qy      451  GAGTTCTTACCAAGGTTGCGGCGCGAGGAAAAGATGTCGCGGTGTCCTGAGAGAT 510
Db      2662 GAGTTCTTACCAAGGTTGCGGCGCGAGGAAAAGATGTCGCGGTGTCCTGAGAGAT 2721
Qy      511  TCCAGTGAACCAATCTTCCACAGTCACTGATGTGAGACAGCTTCCGAGGGAGCGTT 570
Db      2722 TCCAGTGAACCAATCTTCCACAGTCACTGATGTGAGACAGCTTCCGAGGGAGCGTT 2781
Qy      571  GAAACGAGTTCTGAGATCAGAAATGCGCCCTGATCTGACTCTCTTAGGAGAAACATCT 630
Db      2782 GAAACGAGTTCTGAGATCAGAAATGCGCCCTGATCTGACTCTCTTAGGAGAAACATCT 2841
Qy      631  GCCTCTTCTGAAAAGGCAAAAGAGGTGAAAGAGAAAGACACCTCTGACAGTGAAGT 690
Db      2842 GCCTCTTCTGAAAAGGCAAAAGAGGTGAAAGAGAAAGACACCTCTGACAGTGAAGT 2901
Qy      691  GATGCGCTTACCTGTAAGAGACCTTCAGAACCGCCTTCGAGAAAGCGAGAACAACT 750
Db      2902 GATGCGCTTACCTGTAAGAGACCTTCAGAACCGCCTTCGAGAAAGCGAGAACAACT 2961
Qy      751  GTGAGAGAGTCCCTGAGAGGCTGACAGATTCGCTCTGAGAGAAAGCGCAGAGAGAGT 810
Db      2962 GTGAGAGAGTCCCTGAGAGGCTGACAGATTCGCTCTGAGAGAAAGCGCAGAGAGAGT 3021
Qy      811  TCTGCCAAACCTGGAGAGTCCAAATAGGCAAGTCCGAGCAGACAGACACTCTCTTAAG 870
Db      3022 TCTGCCAAACCTGGAGAGTCCAAATAGGCAAGTCCGAGCAGACAGACACTCTCTTAAG 3081
Qy      871  CAGAGAGCTGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAGATGACATAGAAAT 930
Db      3082 CAGAGAGCTGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAGATGACATAGAAAT 3141
Qy      931  CAGTTGAAAGGAGAGCGCACTCAAGGAAATACAGAGAAAACCCAGGAGAGCGGCAAA 990
Db      3142 CAGTTGAAAGGAGAGCGCACTCAAGGAAATACAGAGAAAACCCAGGAGAGCGGCAAA 3201
Qy      991  CCAAGAGCTGAGTGAAGTTTACAGACCCCAATGCGCTGATGACTGCGCGCAGCT 1050
Db      3202 CCAAGAGCTGAGTGAAGTTTACAGACCCCAATGCGCTGATGACTGCGCGCAGCT 3261
Qy      1051 CACAACAACAGGTTTATG 1068
Db      3262 CACAACAACAGGTTATG 3279

```

```

RESULT 7
LOCUS   BUS22670      874 bp      mRNA      linear      EST 13-SEP-2002
DEFINITION  AGNCOURT 10158094 NCI_CGAP Co24 Mus musculus cDNA clone
IMAGE:552926 5', mRNA sequence.
ACCESSION  BUS22670
VERSION    BUS22670.1 GI:22830196
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 874)

```

```

AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
          Email: cgabs-remail.nih.gov
          Tissue Procurement: The Cepko Laboratory
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLNL4129 row: a column: 02
          High quality sequence stop: 720.
          Location/Qualifiers
            1..874
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="FVB/N"
              /db_xref="taxon:10090"
              /clone="IMAG:552926"
              /lab_host="DH10B (71 phage-resistant)"
              /clone_1ib="NCI CGAP Co24"
              /note="Forgan: colon: Vector: pCMV-SPORT6; Site 1: NotI;
              Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
              Average insert size 1.6 kb. Constructed by Life
              Technologies. Note: this is a NCI_CGAP library."

```

ORIGIN

```

Query Match      27.2%  Score 780.8; DB 13; Length 874;
Best Local Similarity 99.5%; Pred. No. 66-180; 2; Indels 2; Gaps 2;
Matches 804; Conservative 0; Mismatches 2;

Qy      797  GCGAGAGAGAGATTCTGCCGAACCTGGAGTGTCCAAATAGGACAGTCCGAGCAGACA 856
Db      1  GCGAGAGAGAGATTCTGCCGAACCTGGAGTGTCCAAATAGGACAGTCCGAGCAGACA 60
Qy      857  GACCTCTCTGTAAGCAGAGAGCTGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAG 916
Db      61  GACCTCTCTGTAAGCAGAGAGCTGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAG 120
Qy      917  ATGACATAGAAATCGATTGGAAGGAGCGCACTCAGGAAATACAGAGAAACCCCA 976
Db      121  ATGACATAGAAATCGATTGGAAGGAGCGCACTCAGGAAATACAGAGAAACCCCA 180
Qy      977  GGGAGAGCGGCAAAACCAAGCCTGAGTGAAGTTTACACCCCAATGCGCTGTACTGCA 1036
Db      181  GGGAGAGCGGCAAAACCAAGCCTGAGTGAAGTTTACACCCCAATGCGCTGTACTGCA 240
Qy      1037  TCTGCCAGAGCTCAACAACAAGTTTATGATCTGTGTATGATGATGATGATGATGAT 1096
Db      241  TCTGCCAGAGCTCAACAACAAGTTTATGATCTGTGTATGATGATGATGATGATGAT 300
Qy      1097  TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1156
Db      301  TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Qy      1157  AAGACTAATGTCGCCCAATTCGACCAATTTGCAAGTGAAGTGAAGACAAACGTTAGCG 1216
Db      361  AAGACTAATGTCGCCCAATTCGACCAATTTGCAAGTGAAGTGAAGACAAACGTTAGCG 420
Qy      1217  CCAACAATGAGAGAGACTGTGGGTGAGATGTGTGGTCTGTATGAGCAAGACTGACAA 1276
Db      421  CCAACAATGAGAGAGACTGTGGGTGAGATGTGTGGTCTGTATGAGCAAGACTGACAA 480
Qy      1277  GCATAGGAGCAGTGAAGCAAGTCCGAGGAGACCAAGGCGTAAAGGGTGAAGATTGAG 1336
Db      481  GCATAGGAGCAGTGAAGCAAGTCCGAGGAGACCAAGGCGTAAAGGGTGAAGATTGAG 540
Qy      1337  AGGAGCAAAACCCAGCGGCAAGAAAACCTCAAGATATTCCAGCTGTGTGAGAGGCTC 1396
Db      541  AGGAGCAAAACCCAGCGGCAAGAAAACCTCAAGATATTCCAGCTGTGTGAGAGGCTC 600

```


ACCESSION CF169555
 VERSION CF169555.1 GI:33279104
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 667)
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
 Genome Res. 11 (9), 1553-1558 (2001)
 JOURNAL MEDLINE
 PUBMED 11544199
 COMMENT Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: B0815 row: C column: 09
 Seq primer: M13 Reverse
 High quality sequence stop: 667
 POLY-A-No.

FEATURES
 source
 1..667
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="taxon:10090"
 /clone="NIA:B0815C09 IMAGE:30468896"
 /dev_stage="Newborn Kidney"
 /lab_host="DH10B"
 /clone_1ib="NIA Mouse Newborn Kidney cDNA Library (long 1)"
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). In brief, double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen): 5'-pAGCTAGTCTAGATCGAGCGGCGCCCTTTTCTTTT-3' from 26 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker l1-SalI, purified by phenol/chloroform, and separated from free linkers by centrifugation 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and restriction 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 Kb. The library was constructed by Yulan Piao."

ORIGIN
 Query Match 23 2%; Score 665.4; DB 14; Length 667;
 Best Local Similarity 99.9%; Pred. No. 1e-151; Indels 0; Gaps 0;
 Matches 666; Conservative 0; Mismatches 1;
 628 CCTGCTCTTCTGAAAGCAAAAGAGGTGAAGAGAAAGACCTCTGACAGTGAC 687
 1 CCTGCTCTTCTGAAAGCAAAAGAGGTGAAGAGAAAGACCTCTGACAGTGAC 60
 688 AGTATGAGCTTACGTTGAAGAACTTCAGAACCCGCTTGGAGAAAGCGAGCGAAGAA 747
 61 AGTATGAGCTTACGTTGAAGAACTTCAGAACCCGCTTGGAGAAAGCGAGCGAAGAA 120
 748 CCTGTGAGAAAGTCCCTGAGAGGAGTCAAGATCGCTGTGAGAAAGCGAGAGGAA 807

Db 121 CCTGTGAGAGAGTCCCTGAGAGGAGTCAAGATCCGCTTGAGAGAAAGCGAGAGAGAA 180
 Oy 808 GATTTCGCCGAAACTGGAGAGTGTCCAAATAGGCAGATCCGACAGAGACGACTCTCT 867
 Db 181 GATTTCGCCGAAACTGGAGAGTGTCCAAATAGGCAGATCCGACAGAGACGACTCTCT 240
 Oy 866 AAGCAGAGAGCTGAGGCTGTGAGGAGCAGAGTGTCCCATGACAGACATGACATAGAA 927
 Db 241 AAGCAGAGAGCTGAGGCTGTGAGGAGCAGAGTGTCCCATGACAGACATGACATAGAA 300
 Oy 928 AATCAGTTGAAAGGAGAGCGACTCAGAGGAAATACAGAGAAACCCCAAGAGGAGG 987
 Db 301 AATCAGTTGAAAGGAGAGCGACTCAGAGGAAATACAGAGAAACCCCAAGAGGAGG 360
 Oy 988 AAACCAAGCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACATGCAATGCCCGC 1047
 Db 361 AAACCAAGCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACATGCAATGCCCGC 420
 Oy 1048 CCTCACAACAGAGGTTTATGATCTGCTGTGATCGGTGTGAGAGAGTGTCCATGAGTAC 1107
 Db 421 CCTCACAACAGAGGTTTATGATCTGCTGTGATCGGTGTGAGAGAGTGTCCATGAGTAC 480
 Oy 1108 TGTGTGGATATTCTGAGGCGCGGCGCTCTCTGAAAGGAAAGCGGAAAGCTACATC 1167
 Db 481 TGTGTGGATATTCTGAGGCGCGGCGCTCTCTGAAAGGAAAGCGGAAAGCTACATC 540
 Oy 1168 TGCCCAATTTGACATTTTGGAGAGTGCAGATGAGCAAAAGGTAGCGGCACATAGAG 1227
 Db 541 TGCCCAATTTGACATTTTGGAGAGTGCAGATGAGCAAAAGGTAGCGGCACATAGAG 600
 Oy 1228 CAGAGCTCTGGGTGAGATCTGTGGGTGCTGATGAGCAGACACTGACAAAGCATAGGACA 1287
 Db 601 CAGAGCTCTGGGTGAGATCTGTGGGTGCTGATGAGCAGACACTGACAAAGCATAGGACA 660
 Oy 1288 GTAGAGC 1294
 Db 661 GTAGAGC 667

RESULT 11
 BB637982 691 bp mRNA linear EST 26-OCT-2001
 LOCUS BB637982
 DEFINITION musculus cDNA clone A630095A07 5', mRNA sequence.
 DBINFORMATION
 BB637982
 BB637982.1 GI:16473713
 EST.
 ACCESSION
 BB637982
 BB637982.1 GI:16473713
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 691)
 Hara, A., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Arakawa, T., Miyazaki, A., Miyazaki, A., Nomura, K., Kono, H., Kondo, M., Koya, S., Matsuyama, T., Sakai, C., Sakai, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Saito, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 Unpublished (2001)
 CONTACT: Yoshitake Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenihiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to

size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCAGCAGC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 21.5%; Score 617.4; DB 13; Length 632;
 Best Local Similarity 99.4%; Pred. No. 5.9e-140;
 Matches 629; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

13 CTGGAATAGCTGCGGATCCCGCGCGCGGAGCTTTTAAAGAGCAGTCCCACTG 72
1 CTGGAATAGCTGCGGATCCCGCGCGCGGAGCTTTTAAAGAGCAGTCCCACTG 60
73 TCCCTTTGGTGTTCGAAGCTCCGGAATTTCTCATTTGAATGAATCTGACCTGAACAGT 132
61 TCCCTTTGGTGTTCGAAGCTCCGGAATTTCTCATTTGAATGAATCTGACCTGAACAGT 120
133 GGGGTGAGGCTTGGCCGTCTGCTGACCTGCCCCAGGCTTAATTTATGTAGGATTT 192
121 GGGGTGAGGCTTGGCCGTCTGCTGACCTGCCCCAGGCTTAATTTATGTAGGATTT 180
193 CCAAGCAAGGTTTTCAGAGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 252
181 CCAAGCAAGGTTTTCAGAGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 240
253 AGCAATGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 312
241 AGCAATGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 300
313 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 372
301 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 360
373 AGTGAAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 432
361 AGTGAAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 420
433 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 492
421 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 480
493 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 552
481 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 540
553 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 612
541 GGTTCGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 600
613 TTAAGGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 645
601 TTAAGGAGAGAGCCCAAGGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCGCACTG 632
  
```

RESULT 13
 CFS35718 623 bp mRNA linear EST 12-SEP-2003
 LOCUS U1-M-GH0-cg2-h-20-0-UI.r1 NIH BMAP GH0 Mus musculus cDNA clone
 DEFINITION IMAGE:30534619 5', mRNA sequence.
 ACCESSION CFS35718
 VERSION CFS35718.1 GI:34587686
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 623)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mouse1.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES

source

Seq primer: pX-5.

Location/Qualifiers

1..623
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30534619"
 /tissue_type="whole brain"
 /dev_stage="1, 5, and 15 days newborn"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP GH0"
 /note="Organ: Brain; Vector: pX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to BMAP, Hemlin and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCAGCAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 21.2%; Score 607.4; DB 14; Length 623;
 Best Local Similarity 99.4%; Pred. No. 1.7e-137;
 Matches 619; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

302 GGAAGAACTGGGGTTTTCGAGAGACCAAGATTGCCAAGCTGAGGCTGAGAGAGACCG 361
1 GGAAGAACTGGGGTTTTCGAGAGACCAAGATTGCCAAGCTGAGGCTGAGAGAGACCG 60
362 AGGCGAGACCCAGT-GAGCAGAGACCAAGAGCATTAACCTCCCTGCGCGCAGTGA 420
61 AGTGAAGACCCAGTGAAGAGAGACCAAGAGCATTAACCTCCCTGCGCGCAGTGA 120
421 CGGCAACCAAAAGTACTGAGAGAGGTTAGAGGTTTTCACACAGGTTGCGCGCAGG 480
421 CGGCAACCAAAAGTACTGAGAGAGGTTAGAGGTTTTCACACAGGTTGCGCGCAGG 180
481 AAAAAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCCAATCTTCCACAGTACT 540
181 AAAAAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCCAATCTTCCACAGTACT 240
541 GATGAGAGAGAGCTTCCGAGGGAGCGTTGAAGAGTTCTGAATCGAAGTGGCCCT 600
241 GATGAGAGAGAGCTTCCGAGGGAGCGTTGAAGAGTTCTGAATCGAAGTGGCCCT 300
601 GATGAGAGAGAGCTTCCGAGGGAGCGTTGAAGAGTTCTGAATCGAAGTGGAA 660
301 GATGAGAGAGAGCTTCCGAGGGAGCGTTGAAGAGTTCTGAATCGAAGTGGAA 360
661 GAGAGAGAGAGAGCTTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 720
361 GAGAGAGAGAGAGCTTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 420
  
```

OY		721	CGCCTTCGGAGAAAGCCGACGACGAAGAACCCTGTGGAGAGCTCCCTGAGAGGCAGTCAAGAT	780
Db		421	CGCCTTCGGAGAAAGCCGACGACGAAGAACCCTGTGGAGAGCTCCCTGAGAGGCAGTCAAGAT	480
OY		781	CGCCTCGAGGAAGAACCGCAGAGAGGAATAATTCTGCCGAACAATCGAGAGTGTCCTAATAGGC	840
Db		481	CGCCTCGAGGAAGAACCGCAGAGAGGAATAATTCTGCCGAACAATCGAGAGTGTCCTAATAGGC	540
OY		841	AGTCCCGAGCAGACAGACACTCTCTGTAAAGCAGAGAGCCTGAGAGCTTAGTACAGAACCATGTG	900
Db		541	AGTCCCGAGCAGCAGACACTCTCTGTAAAGCAGAGAGCCTGAGAGCTTAGTACAGAACCATGTG	600
OY		901	TCCCAGTCGAGAGCAGATGACAT	923
Db		601	TCNCACTGAGAGCAGATGACAT	623
<hr/>				
RESULT 14				
BB636983				
LOCUS				
DEFINITION				
BB636983				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
BB636983 645 bp mRNA linear EST 26-OCT-2000				
BB636983 RIKEN full-length enriched, adult male aorta and vein Mus musculus cDNA clone A530047115 5', mRNA sequence.				
BB636983				
BB636983.1 GI:16472769				
EST.				
Mus musculus (house mouse)				
Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
1 (bases 1 to 645)				
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Iehiji,Y., Ito,M., Kawai,J., Komuro,H., Koude,M., Koye,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,I., Shibata,K., Shingawa,A., Shiraki,F., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.				
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)				
Unpublished (2001)				
Contact: Yoshihide Hayashizaki				
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute				
The Institute of Physical and Chemical Research (RIKEN)				
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
Tel.: 81-45-503-9222				
Fax: 81-45-503-9216				
Email: genome-ree@sc.riken.go.jp,				
URL: http://genome.gsc.riken.go.jp/				
Carninci,P., Shibata,Y., Hayasu,N., Sugahara,Y., Shibata,K., Itoh,M., Komu,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
Normaliztion and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)				
wsgl,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Maezawa,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.				
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)				
Komou,H., Fukuishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.				
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)				
Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamaneke,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.				
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)				
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.				

FEATURES	e mouse tissues.	Location/Qualifiers
source		
	1..645	
	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/db_xref="taxon:10090"	
	/clone="A530047115"	
	/sex="male"	
	/tissue_type="aorta and vein"	
	/dev_stage="adult"	
	/lab_host="DH10B"	
	/clone_1ib="RIKEN full-length enriched, adult male aorta and vein"	
	/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAAGAGAGAGAGATCCAGAGCTCTTTTATTTTATTTN 3'], cDNA was prepared by using Trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAAGAGAGAGATTCGAGATTATTAATTAATTCCTCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX 1."	
ORIGIN		
Query Match	20.7%;	Score 594.6; DB 10; Length 645;
Best Local Similarity	99.2%;	Pred. No. 2.3e-134;
Matches 608; Conservative	0; Mismatches 4; Indels 1; Gaps 1;	
QY	1	TCGGTGTAGCTCTGGAATATGCGCGGAGTCCCGGCGCGGGAGCTTGTAAAGG 60
DB	34	TCCGTGTAGCTCTGGAATATGCGCGGAGTCCCGGCGCGGGAGCTTGTAAAGG 93
QY	61	CAGTCCCACTGTCCCTTGTGTGTGTGAAGCTCCGAATCTTCTCATATGAAGCTGTG 120
DB	94	CAGTCCCACTGTCCCTTGTGTGTGTGAAGCTCCGAATCTTCTCATATGAAGCTGTG 153
QY	121	ACTGACAAACAGTGGGAGTGAAGGCTTGGCGCTGTGCTTGAACCTGACCTGGCCCAAGCTCTAATTT 180
DB	154	ACTGACAAACAGTGGGAGTGAAGGCTTGGCGCTGTGCTTGAACCTGACCTGGCCCAAGCTCTAATTT 213
QY	181	TATGTAGAGATTTCCAGCGCAAGGTTTCCAAAGCTTTCAGTTTGTGGACAGGTATGAGAT 240
DB	214	TATGTAGAGATTTCCAGCGCAAGGTTTCCAAAGCTTTCAGTTTGTGGACAGGTATGAGAT 273
QY	241	AAAGGGCACTGAGCATATGAGAAAGCAACCAAGGCTATCAAAACCACAGTAAGAGATT 300
DB	274	AAAGGGCACTGAGCATATGAGAAAGCAACCAAGGCTATCAAAACCACAGTAAGAGATT 333
QY	301	AGGAAAACCTGAGGATTTTGAAGAACCAAGATTTCCAAACCTGAGGATGACAGACACG 360
DB	334	AGGAAAACCTGAGGATTTTGAAGAACCAAGATTTCCAAACCTGAGGATGACAGACACG 393
QY	361	GAGCGGACCTCCAGTGAAGCAGCAACCAAGCAGATTAACCTTCCCTCCGCGCGGAGTGA 420
DB	394	GAGGTGACCCCAAGTGAAGCAGCAACCAAGCAGATTAACCTTCCCTCCGCGCGGAGTGA 453
QY	421	CGGCAACCAAAAGTACTGAGAGGATGAAGAGTTTCTTACACAGGTTTCGGCGCGGAGG 480
DB	454	CGGCAACCAAAAGTACTGAGAGGATGAAGAGTTTCTTACACAGGTTTCGGCGCGGAGG 513
QY	481	AAAAAGATGTGCGGTGTCCCTGAGAGATTCAGTGAAGCCACATCTTCCAGATCACT 540
DB	514	AAAAAGATGTGCGGTGTCCCTGAGAGATTCAGTGAAGCCACATCTTCCAGATCACT 572
QY	541	GATGTGAGACAGCTTCCAGAGGAGAGGCTTGAAGACAGTTTGAATTCAGAGTGGCCCT 600

Db	Accession	Source	Organism	Reference Authors
Db	573	GAATTCGACAGCTTCCGAGGAGGAGCTTGAAGACAGCTTCTGAGATGAGAGTGGCCCT	632	
Db	601	GTATCTGACTCTCT	613	
Db	633	GTATTGACTCTCT	645	
RESULT 15				
LOCUS	BY752619			
DEFINITION	BY752619 RIKEN full-length cDNA	746 bp	mRNA	linear
ACCESSION	BY752619			
VERSION	BY752619.1			
KEYWORDS	EST			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	1 (bases 1 to 746)			
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nishida, T., Otsu, N., Saito, R., Suzuki, H., Yamana, A., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Kyoshinaka, C., Gotohori, T., Baldarelli, R., Hill, D.E., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chochocka, C., Corbani, L.E., Cousins, S., Datta, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Fraser, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Guetlinich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Kondaya, A., Kurotschin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Mikki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Petre, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sanderlin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Veardo, R., Wagner, L., Walthers, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yangisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayashizumi, N., Hayashizumi, T., Kondo, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., and Hayashizaki, Y.			
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs			
JOURNAL	Nature	420	563-573	(2002)
MEDLINE	22354683			
PIRMEED	12466851			
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-reegsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirokawa, T., Koiwa, Y., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Komori, H., Koyama, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to			

FEATURES

SOURCE

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

Riken integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken. Genomic Sciences Center and Genome Science Laboratory in Riken, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hunaoki Atsugi City Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

ORIGIN

Query Match	20.6%	Score 591.6;	DB 13;	Length 746;
Best Local Similarity	92.7%;	Pred. No. 1.2e-133;		
Matches 651; Conservative	0;	Mismatches 48;	Indels 3;	Gaps 3

[illegible]

Db 317 AGGAAACCTGGGTTTTCGAAAGACGATGCCAAACGTGAGGTCAGAGACAGC 376
Qy 361 GAGGCGAGCCCACTGAGCAGCAACCAAGCAATTAACCTCTCCCTGCGCCGAGTGA 420
Db 377 GAGGTGACCCCACTGAGCAGCAACCAAGCAATTAACCTCTCCCTGCGCCGAGTGA 436
Qy 421 CGGCAACCAAAACCTGAGAGGCTAAGAGTTTCTTACCAAGGTTCCGCGCCGAGG 480
Db 437 CGGCAACCAAAACCTGAGAGGCTAAGAGTTTCTTACCAAGGTTCCGCGCCGAGG 496
Qy 481 AAAAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCCAATCTTCCAGTCACT 540
Db 497 AAAAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCCAATCTTCCAGTCACT 556
Qy 541 GATGTGAGACGCTTCGAGGGAGCGTTGAAAGCAAGTTCTGAGATCAGAAAGTGCCT 600
Db 557 GATGTGAGAGCGCTTCGAGGGAGCG-TGAAAGCAAGTTCTGAAATCAAAAGGCGCCT 615
Qy 601 GTATCTGACTCTTATGGAAGAAACATCTGCTCTTCTGAAAGCAAAAGAGGTGA 660
Db 616 GT-TCTTACTCTTATGAGGAA-AAAATCTGCTTCTTGAAGGCAAGAGGAGGAA 673
Qy 661 GAGGAAGAGACACTCTGACAGTGAAGTATGAGCTTACG 702
Db 674 GAANNAAAACCTTTNAGNGAGAGGAGGCTTACCTTAAG 715

Search completed: April 23, 2004, 18:03:08
Job time : 7086.26 secs


```

;
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: BP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT29pt-Fla
; US-08-232-463-14
;
Query Match          2.6%; Score 67; DB 1; Length 7218;
Best Local Similarity 4.6%; Pred. No. 8,4e-09;
Matches 19; Conservative 235; Mismatches 155; Indels 0; Gaps 0;
;
QY 654 AGCTGTGAAGAACGACGACCTCTTGTGAAAAGTGAAAGAGGAGATGACCAAGATGA 713
    |||||
DB 1464 AGTGTAAAGAGATGAAGAAATTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1405
QY 714 CACCTCCATAGTGAACAGATGCGCTGACCTTGAAAGCTTCAGATCGCTTCGAG 773
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1404 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1345
QY 774 GAAGCGGAACGAGACCTGAGAGCCCTGTAAGAGATGACAGTCCGCTCGGA 833
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1344 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1285
QY 834 GAAGCGCGGAGAGAGGCTCCGCGAGACTGTGGGCTCCGAGCGCACTGACACTGTGA 893
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1284 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1225
QY 894 GGGCGTCTGCGCCATGAAGACGACCGAGAACATACAGGGGGTGTGTCCAGGCTGG 953
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1224 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1165
QY 954 GAAAGATGACAGAGAGTAACTGTGAGGGAAGCGGCTCAGACATCAAGATGAGGA 1013
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1164 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1105
QY 1014 GCCTGGAAGCTTGGCCGACGACGACCTGATATGTAGAGGTTACGACCC 1062
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1104 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1056

```

```

;
; RESULT 3
; US-09-620-312D-269
; Sequence 269, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundl, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Duanli
; APPLICANT: Wang, Zhimei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_fl_genes Version 1.0
; SEQ ID NO 269
; LENGTH: 6256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (257)..(5443)
; US-09-620-312D-269
;
Query Match          2.5%; Score 65; DB 4; Length 6256;
Best Local Similarity 62.7%; Pred. No. 3e-08;
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
;
QY 1060 CCAAGCGCTGTATGATGATTCGCGCAGCTCAACAAACAGGTTATGATTCGT 1119
    |||||
DB 1469 CCAAGCAAGAGTGTGGGTTTGGCAAAAACCAACATGCAAGGTTATGTTGGCTGT 1528
QY 1120 GACGCGTGTGAAGATGTTTCATGCGCATTTGTGGGCAATTCGAGGCTCGAGGAG 1179
    |||||
DB 1529 GGAAGATGTGATGACTGCTTCATGTAATGTGTGTGGTTAAGCTTTCTCAAGCAG 1588
QY 1180 CTTTGAAGGAATGGGGAAGACTATATCTGCCAAACTG 1220
    |||||
DB 1589 CAGATGGCGAGAGAGACAAAGAAATATGTCTGTGAAATG 1629
;
; RESULT 4
; US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHRIFFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria

```

[illegible]

```

Db      706  GACGGGGAGGACACAGAGGACCGGGGAGGAGCGGGGAGGAGACGAGGACCGGAGG  64
Qy      922  GACAGACATATGAGGGGGGTGTGTCCACGGCTTGGGAAAGATGACAGAGAGATTAACTTGGAG  981
Db      646  GACGAGAGACCGGGAGGAGACCGGGGAGAGACGGGAGGAGCGAGGAGCGAGAGGAGGAC  587
Qy      982  GGAAGGCGGCTCAGGACATCAAGATGAGAGGCTTGGAGACTTGGGCGGACCGAAGCTT  1041
Db      586  GGGGAGGAGCGGGAGGAGGACGGGGAGGAGGAGCGAGGAGGAGCGGGGAGGAGGACGAG  527
Qy      1042 GAATCTGAGGCGTTACGA 1058
Db      526  GACGGGAGGAGGAGGACGA 510

RESULT 6
US-09-130-114-2/C
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaji, Baasam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCES: 0867/1D9030US1
; CURRENT APPLICATION NUMBER: US/09/130.114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA

Query Match      1.7%; Score 44.2; DB 2; Length 1931;
Best Local Similarity 44.8%; Pred. No. 0.019;
US-09-130-114-2

```

Matches 169; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 662 GAAAAGTGAAGAGGAGATGACCACTGATGATGACAGGATGCTG 741
DB 886 GACGGGGAGACGGGGAGACGAGAGACGGGGAGACGAGAGCGGGAG 827
QY 742 ACCTTGAAGAGCTTCAAGATCGCTTCCAGAGAGCGGAGACAGAGCCCACTAGAGG 801
DB 826 GACGGGGAGAGACGAGAGACGGGGAGACGAGAGAGACGAGAGAGCGGGAG 767
QY 802 CCCCTGAAGAGATCAGAGTGCCTCGAGAGAAAGCCCGGAGAGAGGCTCCGCGAG 861
DB 766 GACGAGAGACGGGGAGACGGGGAGAGACGAGAGAGAGAGAGAGAGAGAGAG 707
QY 862 ACTGTGGCTCCGAGGCGCAGTGAACCTGTGAGAGGCGCTTCCCACTAGAGAGAGCC 921
DB 706 GACGGGGAGAGACGAGAGACGGGGAGAGACGGGGAGAGAGACGAGAGAGCGGGAG 647
QY 922 GAGAGCATCAGGGGGGTTGTCTCCAGAGCTGGGAAAGATGACAGAGAGTAAATTGAG 981
DB 646 GACGAGAGACGGGGAGAGACGGGGAGAGACGGGGAGAGAGAGAGAGAGAGAGAG 587
QY 982 GAAAAGCGGCTCAGAGACATCAAGATGAGAGAGCTTGGAGACTTGGCCGAGAGGCTT 1041
DB 586 GCGGAGAGACGGGGAGAGACGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 527
QY 1042 GATGTGAGGCTTACGA 1058
DB 526 GACGGGAGAGAGAGCA 510

RESULT 7
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-Ken
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620 Leuburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1642 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-781-891-208

Query Match 1.6%; Score 42.6; DB 3; Length 16442;
Best Local Similarity 48.5%; Pred. No. 0.24;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 774 GAAAGGGAACAGAGAGCCCACTGAGAGAGCCCTGAAAGGATCCAGAGTCCGCGGAA 833
DB 16430 GAG 16371
QY 834 GAAAGGCGGAGAGAGAGGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 893
DB 16370 GAG 16311
QY 894 GAGGCTCTGCTCCCACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 953
DB 16310 GAG 16251
QY 954 GAAAGATGACAGAGAGATGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013
DB 16250 GAG 16191
QY 1014 G 1014
DB 16190 G 16190

RESULT 8
US-09-618-166-208/c
; Sequence 208, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-Ken
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/618,166
; FILING DATE: 17-Jul-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: McMaister, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.419C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-618-166-208

Query Match 1.6%; Score 42.6; DB 4; Length 16442;

Best Local Similarity 48.5%; Pred. No. 0.24;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 774 GAACGGGAAACGAGAGCCCACTGAGAGCCCTTGAAGGATCCGAGATCCGCTGGCGAA 833
Db 16430 GAGACGAG 16371
QY 834 GAAGCGCCGGAG 893
Db 16370 GAG 16311
QY 894 GAGCGCTCTCCCAAG 953
Db 16310 GAG 16251
QY 954 GAAAGATACAGAGAGATTAAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013
Db 16250 GAG 16191
QY 1014 G 1014
Db 16190 G 16190

RESULT 9
US-08-867-030B-5
Sequence 5, Application US/0867030B
Patent No. 5948900
GENERAL INFORMATION:
APPLICANT: Yoher et al.
TITLE OF INVENTION: Streptococcus pneumoniae
TITLE OF INVENTION: Capsular Polysaccharide Genes and Flanking Regions
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Benjamin Aaron Adler, Ph.D., J.D.
STREET: 8011 Candle Lane
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch floppy
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,030B
FILING DATE: June 2, 1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 08/243,546
FILING DATE: May 16, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5923
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 777-2321
TELEFAX: (713) 777-6908
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4951 bp
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: other nucleic acid
HYPOTHEICAL: no
ANTI-SENSE: no
FRAGMENT TYPE:
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
POSITION IN GENOME:

US-08-867-030B-5
Query Match 1.6%; Score 41.8; DB 2; Length 4951;
Best Local Similarity 49.8%; Pred. No. 0.19;
Matches 106; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 2237 ACAACAGCCGCTGTATACGTTTCCAGTATTCACGCGCAGACATGAAAGCA 2296
Db 1238 AGAAAACTTATATACCGGATTTATTTGGCAGAGATTAAGGGGAAATATTTTATGTA 1297
QY 2297 TCACTCTTTTATGTTGAGGAATCTTTCAGATGATGATGATGATGATGATGATGATGAT 2356
Db 1238 TCACTTTATTTTAACTGTTGATTTTTCAGATGATGATGATGATGATGATGATGATGAT 1357
QY 2357 GTACATTTATTTTGTACGCGCAGATTAAGGCAATTTTTCATATGCCATTCAC 2416
Db 1358 GTTTTGTCTTATTTCTATTCGTTGGCGGCTTATATTTTCAATGCTGATGATTA 1417
QY 2417 AGCTTACACCCACCATGACATTCGACTT 2449
Db 1418 GTCTTACAGTTGTAGTATGATGAGAAATT 1450

RESULT 10
PCT-US95-06119-5
Sequence 5, Application PC/TUS9506119
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE CAPSULAR
TITLE OF INVENTION: POLYSACCHARIDE GENES AND FLANKING REGIONS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06119
FILING DATE: CONCURRENTLY HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/243,546
FILING DATE: 16-MAY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: AMCT018P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4951 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: other nucleic acid
HYPOTHEICAL: /desc = "DNA"
PCT-US95-06119-5

Query Match 1.6%; Score 41.8; DB 5; Length 4951;
Best Local Similarity 49.8%; Pred. No. 0.19;
Matches 106; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

2237 ACAACAGCGCTGATACGCTTCAGTTTTCACATGTCACAGCCAGACATGAAGA 2296
1238 AGAAAACTCTATACACCGGATTTATTTGGACAGAAATTAATTTATATGA 1297
2297 TCACCTCTTTTATGTTGGGAATCTTGCAGTATGTTGATCTGATTTTCAAGT 2356
1298 TACATTTATTTAATGTTGATTTTTCAGATCATGATTTTCAATTTCTTATGTT 1357
2357 GTACATTTATTTTACCTGGGAGATAGGGATTTTTCATTCATGTCGATTCAC 2416
1358 GTTTTTCCTTTATCTTATTCGTTGGCGGCTTATATTTTCATGCTGATATATA 1417
2417 AGCTACACACCCACATACACATTCGACTT 2449
1418 GTCTACAGTTGATGTAAGTATGAGAGATT 1450

RESULT 11

US-09-710-794-3
Sequence 3, Application US/09710794
Patent No. 6573069
GENERAL INFORMATION:
APPLICANT: Holloway, James L.
APPLICANT: Gao, Zeren
APPLICANT: Whitmore, Theodore E.
TITLE OF INVENTION: NOVEL CRIB PROTEIN ZMS1
FILE REFERENCE: 99-76
CURRENT APPLICATION NUMBER: US/09/710,794
CURRENT FILING DATE: 2000-11-09
PRIORITY APPLICATION NUMBER: US 60/164,685
PRIORITY FILING DATE: 1999-11-10
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 1068
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence for human zms1
NAME/KEY: m1sc.feature
LOCATION: (1)...(1068)
OTHER INFORMATION: n = A,T,C or G
US-09-710-794-3

Query Match 1.6%; Score 41.4; DB 4; Length 1068;
Best Local Similarity 22.4%; Pred. No. 0.086;
Matches 102; Conservative 79; Mismatches 275; Indels 0; Gaps 0;

513 CCGCGGCGAGAGAGATGCTGTCTCCCTGAGAGATTCGGTGAAGCCACGTCCTGCC 572
102 TMCNCAVATNATGCAVGTGAGMNGCNGANGVGNITTYGAGATACWMSNTTYYTNA 161
573 CGCCACAGACGCCGAGACAGCTCCGAGGGCAGCGCTGAGAAAGCTTCTGAGACCAAG 632
162 YMSNARCGNGNGARCCNGAIVGNGARMSNTTNGAYGARCRCWMSWMSWMSWMS 221
633 CGGCCCCAGTGTCTCCACAGCTGTGAAGAACACAGCCTCTTCTGAAAAGTGA 692
222 NAARMGNSNTTYTWSNMGNAAATTYMNGMNSNAAARMGNSWMSWMSWMSWMS 281
693 AGGAGGGGATGACAGATGACACCTCCGATAGTGAACAGATGAGCCCTGAGAAAG 752
282 NCGNARMNGARCARMGATATGTTNGMNSNTTNGMNGAYWMSWMSWMSWMSWMS 341
753 GCTTCAGATTCCTTCGAGAGAGCGGAGAACAGAGCCCATGAGAGGCCCTGAAAG 812
342 BAAYGNAATGWSNTTNCNCAATYNAAYGABARAGRGNGGAGAAAGNACWMSNA 401
813 GATCCAGATCGCTTCGAGAGAGCGCGGAGAGAGGCTCCGCGGAGACTGTGGCTC 872
402 RYTNCNNAARMSNTTWSWMSWMSWMSWMSWMSWMSWMSWMSWMSWMSWMSWMS 461

873 CGAGCCAGTACACTGTGAGAGGCGCTCTGCTCCAGTAAAGCAGAGCCCGAGACATCA 932
482 YGARGARCGNGNACNGARAGCGNGTNCNMMNGNAAVYGANGCNGANGCNCAYMS 521
933 GGGGTTGTGTCCAGCGCTGGGAAAGATGACAGAG 968
522 NCCNAYCCNTYNTNNGAYGARGARGCTTYYGANGA 557

RESULT 12

US-09-669-751-79
Sequence 79, Application US/09669751
Patent No. 6551575
GENERAL INFORMATION:
APPLICANT: Greenspan, Ralph J.
TITLE OF INVENTION: Methods for Identifying Compounds for
TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to
TITLE OF INVENTION: Balance and the Perception of Gravity
FILE REFERENCE: P-NI 3864
CURRENT APPLICATION NUMBER: US/09/669,751
CURRENT FILING DATE: 2000-09-26
PRIORITY APPLICATION NUMBER: US 60/168,579
PRIORITY FILING DATE: 1999-12-02
NUMBER OF SEQ ID NOS: 261
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 79
LENGTH: 669
TYPE: DNA
ORGANISM: Drosophila
US-09-669-751-79

Query Match 1.6%; Score 41; DB 4; Length 669;
Best Local Similarity 46.1%; Pred. No. 0.082;
Matches 137; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

742 ACCTGAAAAGCTTCAGATGCTTCGACAGAGCCGGAACAGAGCCCATGAGAG 801
201 ACCTGTGACAGAAATTAATGCTGCTCCGCAACCTTTGTGTCTTCATGATGGG 260
802 CCCGTAAGGATGATCAGATGCTGCGGAGAAAGCCCGGAGAGAGGTCCTCCGAG 861
261 GGGGATGAGATTTTCTGAGACATATGAAAGAGAGCCCGGAGAGGCGCGAGAG 320
862 ACTGTGCTCCGAGGCGAGTACACTGTGAGAGGCGTCTCTCCAGTAAAGAGCC 921
321 GCGCGAAATCCCGCAGAGAGAGAGAGAGGCGCAGAGGCGAGAAAGCAGAG 380
922 GAGACGATCAGGGGCTGTGCTCCAGCTGGAAGATGACAGAGAGTAAAGTTGAG 981
381 GAAAAGAGAGGCGGAAAGAGAGAGGCGGAGAGAGAGAGAGAGAGAGAGAG 440
982 GGAAGGCGGCTCAGACATCAAGATGAGAGGCTGAGAGCTTGGCCGACCGAG 1038
441 GGAAGGAAAG 497

RESULT 13

US-09-621-976-17202
Sequence 17202, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTe and Encoded Human Proteins.
FILE REFERENCE: GENSET-054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 17202
LENGTH: 364
TYPE: DNA

ORGANISM: Homo sapiens
US-09-621-976-17202

Query Match 1.6%; Score 40.8; DB 4; Length 364;
Best Local Similarity 13.1%; Pred. No. 0.062;
Matches 36; Conservative 123; Mismatches 115; Indels 0; Gaps 0;

426 GCACGACGTGGGCGCTTCCTGCGCGCCAGTGGAGGAGCCCAAGCCACTGACCGCT 485
10 GRARCCGCCCKGAGTSGKMSKSSRYGRSSSCCGSMGMSGCCSRSRMRCKMSMSMM 69
486 GAGACATTCCTGACCATTTGCGCGCGCCGCGGAGAGAGCATGCTGTCTCCCTGGA 545
70 MYRSMKMYKSTGASCTCKGKMAKMTCTMGYAMRYVMAISYGCYSYARVYTCYSKYM 129
546 GATTCGTGAGACCCCACTGCTCCCGCCCAAGACCCCAAGACGCTCCGAGGCGAG 605
130 WKCYRTRRSGMCMWCMGAGSMCYSRASGRYSKSGRGMWKKGSRATSKKGRMMW 189
606 CGTGAAGAAGCGCTCTGAGACCAAGAGCGCGCCCAAGTCTCTTCCACAGCTGTGAAGA 665
190 MKGSRBRATSRVGMMSMYASIRMSKCSASTRMSASCMYTMMSAGSYASGAMWS 249
666 ACGACGACCTCTTCTGAAAAGTGAAGAGAGG 699
250 KYRCATKMSCTYSWYMRASMKSKYCAWSRSGSK 283

RESULT 14

US-09-007-005-17
Sequence 17, Application US/09007005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihne
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
TITLE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007.005B
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035.963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064.491
EARLIER FILING DATE: 1997-11-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Translation template
NAME/KEY: misc feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.5%; Score 40.2; DB 3; Length 289;
Best Local Similarity 5.6%; Pred. No. 0.079;
Matches 12; Conservative 97; Mismatches 106; Indels 0; Gaps 0;

808 AAAGGATCCAGAGTCGCTGCGGAAGAGCCCGGAGAGAGGCTCCGCGCAAGCTGTG 867
51 RARURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 110
868 GGCTCCGAGCAGTGAAGTGTGAGGAGGCTCTGCGCAAGTGAAGAGAGCCGAGAAC 927
111 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 170
928 GATCAGGGGGTGTGTCCAGGCTGGAAAGTGAAGAGAGAGTAACTTGAGGGAAG 987

DB 171 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 230
QY 988 GCGGCTCAGACATCAAGATGAGAGGCTTGAGA 1022
DB 231 RCRGRURARCRURCRURURGRGRCGRUAAAAA 265

RESULT 15

US-09-244-796-17
Sequence 17, Application US/09244796
Patent No. 6281344
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihne
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
TITLE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350007
CURRENT APPLICATION NUMBER: US/09/244.796
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/035.963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064.491
EARLIER FILING DATE: 1997-11-06
EARLIER APPLICATION NUMBER: 09/007.005
EARLIER FILING DATE: 1998-01-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Translation template
NAME/KEY: misc feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.5%; Score 40.2; DB 3; Length 289;
Best Local Similarity 5.6%; Pred. No. 0.079;
Matches 12; Conservative 97; Mismatches 106; Indels 0; Gaps 0;

808 AAAGGATCCAGAGTCGCTGCGGAAGAGCCCGGAGAGAGGCTCCGCGCAAGCTGTG 867
51 RARURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 110
868 GGCTCCGAGCAGTGAAGTGTGAGGAGGCTCTGCGCAAGTGAAGAGAGCCGAGAAC 927
111 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 170
928 GATCAGGGGGTGTGTCCAGGCTGGAAAGTGAAGAGAGAGTAACTTGAGGGAAG 987
171 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 230
QY 988 GCGGCTCAGACATCAAGATGAGAGGCTTGAGA 1022
DB 231 RCRGRURARCRURCRURURGRGRCGRUAAAAA 265

Search completed: April 23, 2004, 18:09:01
Job time: 191.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 14:17:01 ; Search time 1241.38 Seconds
(without alignments)
9479.366 Million cell updates/sec

Title: US-09-787-016a-1
Sequence: 1 ctccgtgcgcgcgcac.....actcttaagatcatcatctg 2610

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues
Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing filter 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptcodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptcodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptcodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptcodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptcodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptcodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptcodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptcodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptcodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptcodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptcodata/1/pubpna/US09_NEW_PUB.seq:*
- 12: /cgn2_6/ptcodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptcodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptcodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptcodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptcodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptcodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptcodata/1/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptcodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1564.4	59.9	1764	9	US-09-925-301-327 Sequence 327, App
2	645	24.7	645	15	US-10-029-386-25091 Sequence 25091, A
3	435	16.7	478	10	US-09-918-995-19127 Sequence 19127, A
4	152.6	5.8	536	15	US-10-029-386-11381 Sequence 11381, A
5	91.2	3.5	226	16	US-10-131-827-8445 Sequence 8445, A
6	65	2.5	5975	10	US-09-873-367C-112 Sequence 112, App
7	65	2.5	6256	15	US-10-117-722-269 Sequence 269, App
8	65	2.5	6256	16	US-10-037-270-269 Sequence 269, App
9	65	2.5	6948	10	US-09-930-213-315 Sequence 315, App
10	60	2.3	60	10	US-09-908-975-6269 Sequence 6269, App
11	52.8	2.0	469	13	US-10-085-783A-56748 Sequence 56748, A
12	52.8	2.0	469	16	US-10-242-535A-56748 Sequence 56748, A
13	52.8	2.0	483	9	US-09-864-761-2140 Sequence 2140, App
14	51.4	2.0	915	9	US-09-764-864-471 Sequence 471, App

15	51.4	2.0	2320	15	US-10-175-523-142	Sequence 142, App
16	50	1.9	50	16	US-10-131-827-916	Sequence 916, App
17	48	1.8	594	13	US-10-142-426-10	Sequence 10, App1
18	48	1.8	594	15	US-10-123-155-10	Sequence 10, App1
19	48	1.8	594	15	US-10-146-731-10	Sequence 10, App1
20	48	1.8	594	15	US-10-140-472-10	Sequence 10, App1
21	48	1.8	594	15	US-10-141-761-10	Sequence 10, App1
22	48	1.8	594	15	US-10-142-885-10	Sequence 10, App1
23	48	1.8	594	15	US-10-158-790-10	Sequence 10, App1
24	48	1.8	594	16	US-10-137-871-10	Sequence 10, App1
25	48	1.8	594	16	US-10-140-923-10	Sequence 10, App1
26	48	1.8	594	16	US-10-141-756-10	Sequence 10, App1
27	48	1.8	594	16	US-10-141-759-10	Sequence 10, App1
28	48	1.8	594	16	US-10-140-805-10	Sequence 10, App1
29	48	1.8	594	16	US-10-140-864-10	Sequence 10, App1
30	44.4	1.7	7386	10	US-09-819-104A-6	Sequence 6, App1
31	44.4	1.7	7534	13	US-10-087-192-650	Sequence 650, App
32	44.4	1.7	8544	10	US-09-819-104A-4	Sequence 4, App1
33	44.4	1.7	91141	13	US-10-087-192-649	Sequence 649, App
34	42.6	1.6	16442	16	US-10-374-077-208	Sequence 208, App
35	42	1.6	2085	17	US-10-311-623-18	Sequence 18, App1
36	42	1.6	2771	13	US-10-112-944-494	Sequence 494, App
37	42	1.6	3195	15	US-10-112-944-494	Sequence 143, App
38	42	1.6	3195	15	US-10-205-823-143	Sequence 54, App1
39	42	1.6	3195	16	US-10-295-976-54	Sequence 1127, App
40	42	1.6	3206	13	US-10-112-944-6	Sequence 6, App1
41	42	1.6	1223197	13	US-10-027-632-179264	Sequence 179264, App
42	42	1.6	1223197	16	US-10-027-632-179264	Sequence 640, App
43	41.8	1.6	636	9	US-09-770-149-640	Sequence 20, App1
44	41.8	1.6	659158	9	US-09-771-208-20	Sequence 2317, App
45	41.2	1.6	4056	15	US-10-156-761-2317	

ALIGNMENTS

RESULT 1
US-09-925-301-327
Sequence 327, Application US/09925301
Patent No. US2002052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 327
LENGTH: 1764
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1758)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1759)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1762)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-327
Query Match 59.9%; Score 1564.4; DB 9; length 1764;
Best Local Similarity 99.6%; Pred. No. 0;

Matches	1610;	Conservative	0;	Mismatches	2;	Indels	5;	Gaps	4;
Qy	996	GGACATCAAAAGATGAGAGCTTGAGAC- TTGGGCCGACCGAAGCTTGATGTGAGGTT							1054
Db	1	GGACATCAAAAGATGAGAGCTTGAGAGCTTTGGGGCGAACCGAAGCTTAATGTGAGGTT							60
Qy	1055	ACGACCCCAAGGCGCTGATGATTTGGCCGACGCTCAAGCAAGAGGTTTATGATT							1114
Db	61	ACGACCCCAAGGCGCTGATGATTTGGCCGACGCTCAAGCAAGAGGTTTATGATT							120
Qy	1115	GCTGTGACCGCTGTGAGAGATGTTTCATGCGGATTTGTGTGGGCAATTTCTGAGGCTCGAG							1174
Db	121	GCTGTGACCGCTGTGAGAGATGTTTCATGCGGATTTGTGTGGGCAATTTCTGAGGCTCGAG							180
Qy	1175	GGAGGCTTTTGGAAAGAAATGGGGAAAGCTATATCTGCCAACTGCAACATTTCTGCAAG							1234
Db	181	GGAGGCTTTTGGAAAGAAATGGGGAAAGCTATATCTGCCAACTGCAACATTTCTGCAAG							240
Qy	1235	TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAGCTAAATGAGACCTGGAG							1294
Db	241	TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAGCTAAATGAGACCTGGAG							300
Qy	1295	ATGCTGATGGACCCGATTTGTACAGATATAGGAACATATAGAGAGAACTTACGCAAGAC							1354
Db	301	ATGCTGATGGACCCGATTTGTACAGATATAGGAACATATAGAGAGAACTTACGCAAGAC							360
Qy	1355	AAGGATTAAGGATTTGAATTTGAGAAAGCTGCAATTCGAATGAGCAAGAAAGAACTCAGA							1414
Db	361	AAGGATTAAGGATTTGAATTTGAGAAAGCTGCAATTCGAATGAGCAAGAAAGAACTCAGA							420
Qy	1415	TCTTCAGACCTGTGTATAGAGGCGCTGGTGCCTCAAAATGTATTTGGCCCGGGCTGTGC							1474
Db	421	TCTTCAGACCTGTGTATAGAGGCGCTGGTGCCTCAAAATGTATTTGGCCCGGGCTGTGC							480
Qy	1475	ACGTGGCGCAGCCCGCATCTGGTGTATCTGCAATGATGACTGTATCTCAACACGCGCAG							1534
Db	481	ACGTGGCGCAGCCCGCATCTGGTGTATCTGCAATGATGACTGTATCTCTCAACACGCGCAG							539
Qy	1535	CGACATTAAGTTTCTTAAGTCTAGGTAAAGAACAGAACCTTAAGGAAAGAAAGATGA							1594
Db	540	CGACATTAAGTTTCTTAAGTCTAGGTAAAGAACAGAACCTTAAGGAAAGAAAGATGA							599
Qy	1595	AGATGAAGCAGAGAAAGCCAGTCTTCGAAATGCGGATCTCAGGAGGATTTAAATCT							1654
Db	600	AGATGAAGCAGAGAAAGCCAGTCTTCGAAATGCGGATCTCAGGAGGATTTAAATCT							659
Qy	1655	CTTCTGTGCAAGAGACCAAGCTCCAGAAAAAAGAGAACCAAGTGAAGAGGACGTGG							1714
Db	660	CTTCTGTGCAAGAGACCAAGCTCCAGAAAAAAGAGAACCAAGTGAAGAGGACGTGG							719
Qy	1715	TGGTCCCTGCGCGAGTGAAGCACTCGGAAAGAAAGCAAGCTTTGTAGACAGACGCCGT							1774
Db	720	TGGTCCCTGCGCGAGTGAAGCACTCGGAAAGAAAGCAAGCTTTGTAGACAGACGCCGT							779
Qy	1775	CGTGGGCGAGGATCAACAATTACATGCAATTAAGCGAAAGAAAGCTGCTCCCTGCG							1834
Db	780	CGTGGGCGAGGATCAACAATTACATGCAATTAAGCGAAAGAAAGCTGCTCCCTGCG							839
Qy	1835	CGTCACTGTTTATTAATGTATGTATCACTAGGGGTTGGCTCTCTGAGACCCCTCCGTT							1894
Db	840	CGTCACTGTTTATTAATGTATGTATCACTAGGGGTTGGCTCTCTGAGACCCCTCCGTT							899
Qy	1895	CTTCTGTGATGAGCAATCCCTGCGGCTGTCCAGGACTTGGAGTTTGACACTTTGTGTTAAG							1954
Db	900	CTTCTGTGATGAGCAATCCCTGCGGCTGTCCAGGACTTGGAGTTTGACACTTTGTGTTAAG							959
Qy	1955	CTGATCAAGACACCGGCTGACCATCAGCGGAAAGAGAGCCCATGTCAGAGATGCTTC							2014
Db	960	CTGATCAAGACACCGGCTGACCATCAGCGGAAAGAGAGCCCATGTCAGAGATGCTTC							1019
Qy	2015	CTGCTGCCCTGTGTCCATCTCTAGTCTGTCAAGACTTCTCTCACTGTTTCCAAAGCTG							2074
Db	1020	CTGCTGCCCTGTGTCCATCTCTAGTCTGTCAAGACTTCTCTCACTGTTTCCAAAGCTG							1079

Qy	2075	TAAACCTACTGTGTAAGCTTACCTTAATGATTGATTCTTAAATCTGTTTCACTCT							2134
Db	1080	TAAACCTACTGTGTAAGCTTACCTTAATGATTGATTCTTAAATCTGTTTCACTCT							1139
Qy	2135	CAGGCTGTGTAAGTATTTGATTTCTTTCAATCCAGTCTGATTTGATAGCAGACTGCC							2194
Db	1140	CAGGCTGTGTAAGTATTTGATTTCTTTCAATCCAGTCTGATTTGATAGCAGACTGCC							1199
Qy	2195	CGGACGCAATCCACCCCTGTCTGCAATGAGTTTGTGACAAAGCGCTGTATAG							2254
Db	1200	CGGACGCAATCCACCCCTGTCTGCAATGAGTTTGTGACAAAGCGCTGTATAG							1259
Qy	2255	CTTCAGTTTTCACATTTGTCACAGGCGAGACATGAAGATGACTCTTTTATGTT							2314
Db	1260	CTTCAGTTTTCACATTTGTCACAGGCGAGACATGAAGATGACTCTTTTATGTT							1319
Qy	2315	GTGGGAATCTTTCAGATTTGATGATCTGATCTGATTTTCAAGGTGATCAATTTATTTGACT							2374
Db	1320	GTGGGAATCTTTCAGATTTGATGATCTGATCTGATTTTCAAGGTGATCAATTTATTTGACT							1379
Qy	2375	GGGAGATAGGGGATTTTTTTTTTTTTCATGTCGATTCACAGCTTACACCCACATG							2434
Db	1380	GGGAGATAGGGGATTTTTTTTTTTTTCATGTCGATTCACAGCTTACACCCACATG							1437
Qy	2435	AACCATTCGAATCTTGAAAGG-CAACACCTGCTGATGAGGCGCCAGGTAAGT							2493
Db	1438	AACCATTCGAATCTTGAAAGG-CAACACCTGCTGATGAGGCGCCAGGTAAGT							1497
Qy	2494	TCACACCTAGAACCTGTCTGACCGCAGAGACGCTGCTTGATGAGCTTGATTTCAATG							2553
Db	1498	TCACACCTAGAACCTGTCTGACCGCAGAGACGCTGCTTGATGAGCTTGATTTCAATG							1557
Qy	2554	TGACTGCTTTCTTGGCTCTGCTCTTTGAATTTAGCTCTTAAGATCATCTCTG							2610
Db	1558	TGACTGCTTTCTTGGCTCTGCTCTTTGAATTTAGCTCTTAAGATCATCTCTG							1614

RESULT 2

US-10-029-386-25091/c
 ; Sequence 25091, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pem, Sharon G.
 ; APPLICANT: Rank, David R.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
 ; FILE REFERENCE: AEMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 25091
 ; LENGTH: 645
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURES:
 ; OTHER INFORMATION: MAP TO AL117379.14
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
 ; OTHER INFORMATION: SWISSPROT HIT: P23583, EVALU2 2.50e-01
 ; OTHER INFORMATION: EST HUMAN HIT: AL580659.1, EVALU0 0.00e+00
 ; OTHER INFORMATION: NT HIT: g16193859, EVALU0 0.00e+00
 US-10-029-386-25091

Query Match 24.7%; Score 645; DB 15; Length 645;
 Best Local Similarity 100.0%; Pred. No. 1.5e-196;
 Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	263	GTATGAGACGAAAGGCGACCCGAGCAATGAGAGGCACTTAAGGCTTCAAAACCAACA							322
Db	645	GTATGAGACGAAAGGCGACCCGAGCAATGAGAGGCACTTAAGGCTTCAAAACCAACA							586

323 GCAAGAGTTCAGGAAAAATGCGGTTTTCGAGAGACATATGCGCAAGCGAGAGCGCG 382
585 GCAAGAGTTCAGGAAAAATGCGGTTTTCGAGAGACATATGCGCAAGCGAGAGCGCG 526
383 CAGGGAGCGCGAGGCTGAGACCTGAGAGCCGCCACAGAGAGAGCTGGGCTGT 442
525 CAGGGAGCGCGAGGCTGAGACCTGAGAGCCGCCACAGAGAGAGCTGGGCTGT 466
443 CCGTGGCGCGAGGCTGAGAGCCGCCACAGAGAGAGCTGGGCTGTCTGACCA 502
465 CCGTGGCGCGAGGCTGAGAGCCGCCACAGAGAGAGCTGGGCTGTCTGACCA 406
503 TTGGCGCGCGCGCGCGAGAGAGACATGCTGTCTCTGAGAGATTTCTGAGAGCCCA 562
405 TTGGCGCGCGCGCGCGAGAGAGACATGCTGTCTCTGAGAGATTTCTGAGAGCCCA 346
563 CGTCTGCGCGCGCGCGAGAGAGACATGCTGTCTCTGAGAGAGAGCTGGGCTGT 622
345 CGTCTGCGCGCGCGCGAGAGAGACATGCTGTCTCTGAGAGAGAGCTGGGCTGT 286
623 AGAGCAGAGAGCGCGCGCGAGAGAGACATGCTGTCTCTGAGAGAGAGCTGGGCTGT 682
285 AGAGCAGAGAGCGCGCGCGAGAGAGACATGCTGTCTCTGAGAGAGAGCTGGGCTGT 226
683 AAAAGGTGAAAG 742
225 AAAAGGTGAAAG 166
743 CTTTGAAGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
165 CTTTGAAGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 106
803 CCGTGAAG 862
105 CCGTGAAG 46
863 CTTGAGGCTCCGAG 907
45 CTTGAGGCTCCGAG 1

RESULT 3
US-09-918-995-19127
; Sequence 19127, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19127
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) - (478)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-19127

Query Match 16.7%; Score 435; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.8e-129;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 905 CCAAGTAAAGCAG 964
44 CCAAGTAAAGCAG 103

QY 965 GAGAGAGTAAAGTGAAG 1024
DB 104 GAGAGAGTAAAGTGAAG 163
QY 1025 TGGGCGCGAG 1084
DB 164 TGGGCGCGAG 223
QY 1085 GCCAGCTCACAAG 1144
DB 224 GCCAGCTCACAAG 283
QY 1145 GCGATGTTGTTGAG 1204
DB 284 GCGATGTTGTTGAG 343
QY 1205 ATATCTGCGCGAG 1264
DB 344 ATATCTGCGCGAG 403
QY 1265 ATCAGCAG 1324
DB 404 ATCAGCAG 463
QY 1325 GAACATAG 1384
DB 464 GAACATAG 1443

RESULT 4
US-10-029-386-11381/c
; Sequence 11381, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: ABOVICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11381
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL117379.14
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P80398, EVALU8 3.80e+00
; OTHER INFORMATION: NT HIT: gi14769835, EVALU8 1.00e-78
; OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALU8 2.00e-78
US-10-029-386-11381

Query Match 5.8%; Score 152.6; DB 15; Length 536;
Best Local Similarity 94.6%; Pred. No. 9.5e-38;
Matches 158; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 244 GCTTACGTTGAG 303
DB 167 GCTTACGTTTACGCTAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 108
QY 304 AAGGCAATCAAAAG 363
DB 107 AAGGCAATCAAAAG 48
QY 364 ATGCGCAAG 410
DB 47 ATGCGCAAG 1

```
RESULT 5
US-10-131-827-8445/c
; Sequence 8445, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
; APPLICANT: Mohlemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 506612000120
; CURRENT APPLICATION NUMBER: US/10/131,827
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8445
; LENGTH: 226
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-8445

Query Match      3.5%; Score 91.2; DB 16; Length 226;
Best Local Similarity 72.5%; Pred. No. 3.4e-18;
Matches 145; Conservative 0; Mismatches 53; Indels 2; Gaps 2;

QY 2153 TGTATCTCTTCATCCGATGTCGATTCGATGACCACTGCGGACAGCCCAATCCACC 2212
    |||||
DB 217 TTTATTTTTCATCCCGTGTGAGTTTCAGATTCATGCTCGCTCAACATCCACT 158
    |||||

QY 2213 CCTGCTGCACATGATGTTCTGCAACAAGCGCTGTATACGCTTCAG-TTTTCCACAT 2271
    |||||
DB 157 CCAGTTTCACATGACATGTTTTCATGTCATGTCATGTCATGTTTTCATGTTTTCAT 98
    |||||

QY 2272 TGTCCAGCGCCGACACATGAAAGCAT-CACTTCTTTTATGTTGTGGATCTTTGCA 2330
    |||||
DB 97 AGTCCACTGCCGAGACATTAATGATATGTTTATGTTGTGTAATCTTGCGCA 38
    |||||

QY 2331 GTTAGTGTTCATCTGATTT 2350
    |||||
DB 37 GTTAGTATAGCATACATATT 18
    |||||

RESULT 6
US-09-873-367C-112
; Sequence 112, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Andrews, Gregory
; APPLICANT: Augustus, Meena
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; TITLE OF INVENTION: Signature Gene Sets
; FILE REFERENCE: 689290-64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 1067
```

```
SOFTWARE: PatentIn version 3.0
; SEQ ID NO 112
; LENGTH: 5975
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-112

Query Match      2.5%; Score 65; DB 10; Length 5975;
Best Local Similarity 62.7%; Pred. No. 9.1e-09;
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1060 CCACAAGCCCTGATATGATTTGCGGACGCTCAACAACAAGTTATGATTTGCT 1119
    |||||
DB 1200 CCGAGACAGAGTGTGGTTTGGAAAAACAATGGCAAGGTTATGTTGGCTGT 1259
    |||||

QY 1120 GACCGCTGTGAAGAATGTTTTCATGCGCATTTGTGTGGCATTTCTGAGGCTGAGGAG 1179
    |||||
DB 1260 GCGAGATGTGATGACTGTGTTTCATGTTGATTTGTGTGTTAAGTTTCTCAACGACAG 1319
    |||||

QY 1180 CTTTGGAAAGGAATGGGGAAGACTATATCTGCCAATCTG 1220
    |||||
DB 1320 CAGATGGCGGAGAGACAAAGAAATATGCTGTGTAATG 1360
    |||||

RESULT 7
US-10-037-270-269
; Sequence 269, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghaast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784C1P2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pc_FL_genes Version 1.0
; SEQ ID NO 269
; LENGTH: 6256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (257)..(5443)
US-10-037-270-269

Query Match      2.5%; Score 65; DB 15; Length 6256;
Best Local Similarity 62.7%; Pred. No. 9.4e-09;
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1060 CCCAAGCCCTGATATGATTTGCGGACGCTCAACAACAAGTTATGATTTGCT 1119
    |||||
DB 1469 CCGAGACAGAGTGTGGTTTGGAAAAACAATGGCAAGGTTATGTTGGCTGT 1528
    |||||
```

QY	1120	GAACGCTGGAAGAAATGGTTTCAATGGGATGTGTGGCAATTCTGAGGCGCTGAGGGAG	1179
Db	1529	GGAGATGTATATACCTGCTTCACTGCATGTGTGTGGCTTAAGTCTTTCATGACACAG	1588
QY	1180	CTTTTGGAAAGAAATGGGGAAGACTATATCTGCCCAACTG	1220
Db	1589	CAGATGGCGGAGACCAAGAAATATGTCTGTGTAAATG	1629

RESULT 8
119-10-11

Sequence 269, Application US/10117722
Publication No. US2003021974A1

```

1 GENERAL INFORMATION:
2 APPLICANT: Tang, Y. Tom
3 APPLICANT: Liu, Chenghua
4 APPLICANT: Aoudi, Vinod
5 APPLICANT: Zhang, Jie
6 APPLICANT: Dymnac, Radoje T.
7 TITLE OF INVENTION: No. US2003021974A1el Nucleic Acids and
8 TITLE OF INVENTION: Polypeptides
9 FILE REFERENCE: 784CIP2BCIP
10 CURRENT APPLICATION NUMBER: US/10/117,722
11 PRIOR FILING DATE: 2002-04-04
12 PRIOR APPLICATION NUMBER: 09/620,312
13 PRIOR FILING DATE: 2000-07-19
14 PRIOR APPLICATION NUMBER: 09/552,317
15 PRIOR FILING DATE: 2000-04-25
16 PRIOR APPLICATION NUMBER: 09/468,725
17 PRIOR FILING DATE: 2000-01-21
18 NUMBER OF SEQ ID NOS: 1104
19 SOFTWARE: pc_Fl_genes Version 1.0
20 SEQ ID NO 269
21
22 LENGTH: 6256
23
24 TYPE: DNA
25
26 ORGANISM: Homo sapiens
27
28 FEATURE:
29 NAME/KEY: CDS
30 LOCATION: (257)..(5443)
31
32 OS-10-117-722-269

```

Query Match Similarity	2.5%	Score 65	DB 16	length 6256
Best Local Similarity	62.7%	Pred. No.	9.4e-09	
Matches 101, Conservative	0	Mismatches	60	Indels 0
				Gaps 0

QY	DB	QY	DB	QY	DB	QY	DB
1060	CCCCAGCCCTGATATTCATTTGCGCGCAGGCTCAACAACAGATTATGATTTGCTGT	1119					
1469	CCGACGACAGCAGTGTGGTTTTTCGAAAAACCATGCGAACAGATTATGTTGGCTGT	1528					
1120	GACCGCTGTGAAGAATGTTTCATGCGCATTTGTGTGGCATTTCTGAGGCTCGAGGAGG	1179					
1529	GGGAGATGATATGACTGTTTCATGAGTATGTGTGGTTAAAGTCCTTCTCAAGCAG	1588					
1180	CTTTTGAAAGCAATGGGGAAGACTATATCTGCCCAACTG	1220					
1589	CAGATGGCGAGGAGCAAAAGATATGTCGTGTAATG	1629					

RESULT 9
US-09-93

; Sequence 315, Application US/09930213
; Publication No. US20030170625A1

GENERAL INFORMATION:
 APPLICANT: ROSENTHAL, ANDRE
 APPLICANT: HINZMANN, BEEND
 APPLICANT: SCHAEFER, REINHARD
 APPLICANT: ZUBER, JOHANNES
 APPLICANT: TCHÉ-NITES, OLEG
 APPLICANT: GRIPS, MARTIN
 APPLICANT: HELLMGEL, MARTIN
 APPLICANT: SCHMITZ, ANNE-CHANTAL
 APPLICANT: SERS, CHRISTINE

```

; TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
;
; FILE REFERENCE: ALBER-14
;
; CURRENT APPLICATION NUMBER: US/09/930,213
;
; CURRENT FILING DATE: 2001-01-31
;
; PRIOR APPLICATION NUMBER: DE 10004102.7
;
; PRIOR FILING DATE: 2000-01-31
;
; NUMBER OF SEQ ID NOS: 885
;
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 315
;
; LENGTH: 6948
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
US-09-930-213-315

```

Query Match	2.5%	Score 65;	DB 10;	Length 6948;
Best Local Similarity	62.7%	Pred. No. 1e-08;		
Matches 101; Conservative	0;	Mismatches 60;	Indels 0;	Gaps 0;

QY	Db	QY	Db	QY	Db
1060	CCCAAGCCCTGATTTATTCATTTCGCCACGCTCACAACAAGATTATGATTTGCTGT	1119	1119	1180	CTTTTGAAAGAAATGAGGAACTATATCTGCGCAAACTG
1119	1119	1119	1119	1180	1180
2173	CCGAGCAAGCAAGTGTGGGTTTTCGAAAAAACCAATGGCAACAGGTTTAGTGTTGGCTGT	2232	2232	2293	CAGATGGGCGAGGAAAGACAAAGAAATATGTCGTGTTAAATG
2232	2232	2293	2293	2293	2293
1120	GACCGCTGTGAAGATATGTTTCATGGCGATTGTGTGGGCACTTTCTAGAGCTCCAGGAGG	1179	1179	2233	GGGAGATGTGATACACGTTCATGTGTGATGTGTGGGTTAAGTCCTTCTCAAGCAG
1179	1179	2233	2233	2233	2233

RESULT 10
US-09-908

; Sequence 6269, Application US/09908975
; Publication No. US20030165843A1

```

1  GENERAL INFORMATION:
2  APPLICANT: SHOSHAN, Avi
3  APPLICANT: VASSERMAN, Alon
4  APPLICANT: MINTZ, Eli
5  APPLICANT: MINTZ, Liat
6  APPLICANT: FAIGLER, Simchon
7  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE V
8  TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
9  FILE REFERENCE: 36688-0005
10 CURRENT APPLICATION NUMBER: US/09/908,975
11 CURRENT FILING DATE: 2001-07-20
12 PRIOR APPLICATION NUMBER: US 60/287,724
13 PRIOR FILING DATE: 2001-05-02
14 PRIOR APPLICATION NUMBER: US 60/221,607
15 PRIOR FILING DATE: 2000-07-28
16 NUMBER OF SEQ ID NOS: 32337
17 SOFTWARE: PatentIn version 3.0
18 SEQ ID NO 6269
19 LENGTH: 60
20 TYPE: DNA
21 ORGANISM: Homo sapiens
22 US-09-908-975-6269

```

Query Match	2.3%	Score 60;	DB 10;	Length 60;
Best Local Similarity	100.0%	Pred. No. 1.6e-08;		
Matches 60;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

1220 GCAACATTCGCAAGTCGACGATGAGACTCATTGAGAAACGGACATCAGCAGGAAGCTA 1279
1 GACACATTCGCAAGTCGACGATGAGACTCATTGAGAAACGGACATCAGCAGGAAGCTA 60

RESULT 11
ITS-10-085

```

: Sequence 56748, Application US/10085783A
: Publication No. US20040037841A1
: GENERAL INFORMATION:
: APPLICANT: ChondroGene Inc.

```

APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 56748
LENGTH: 469
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5)..(5)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (30)..(30)
OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-56748

Query Match 2.0%; Score 52.8; DB 13; Length 469;
Best Local Similarity 65.0%; Pred. No. 1.4e-05;
Matches 78; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1101 CAGGTTATGATTGCTGTCGACCGCTGTGAGAGATGTTTCATGCGGATTTGGGCGAT 1160
DB 116 CAGGTTATGATTGCTGTCGACCGCTGTGAGAGATGTTTCATGCGGATTTGGGCGAT 1175
QY 1161 TTCTGAGCTCGAGGAGCTTTTGGAAAGATGGGAGACTATCTGTCGCCAAACTG 1220
DB 176 AAGTCTTCTCAAGCAGCAGATGCGCGAGAGCAAGATATGCTGTGTAATG 235

RESULT 12
US-10-242-535A-56748
Sequence 56748, Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 56748
LENGTH: 469
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5)..(5)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (30)..(30)
OTHER INFORMATION: n is a, c, g, or t

US-10-242-535A-56748

Query Match 2.0%; Score 52.8; DB 16; Length 469;
Best Local Similarity 65.0%; Pred. No. 1.4e-05;
Matches 78; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1101 CAGGTTATGATTGCTGTCGACCGCTGTGAGAGATGTTTCATGCGGATTTGGGCGAT 1160
DB 116 CAGGTTATGATTGCTGTCGACCGCTGTGAGAGATGTTTCATGCGGATTTGGGCGAT 1175
QY 1161 TTCTGAGCTCGAGGAGCTTTTGGAAAGATGGGAGACTATCTGTCGCCAAACTG 1220
DB 176 AAGTCTTCTCAAGCAGCAGATGCGCGAGAGCAAGATATGCTGTGTAATG 235

RESULT 13
US-09-864-761-2140
Sequence 2140, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2140
LENGTH: 483
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL050329.9

```

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
US-09-864-761-2140

Query Match      2.0%; Score 52.8; DB 9; Length 483;
Best Local Similarity 65.0%; Pred. No. 1.4e-05;
Matches 78; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 1101 CAGTTTATGATTGCTGTGACCGCTGTGAAGATGTTTCATGCGCATTTGTGGGCA 1160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 CAGTTTATGTTGGCTGTGGAGATGTGATGACTGGTTTCATGTTGATTTGTGGGTT 278
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1161 TTCTGAGGCTGAGGAGGCTTTTGAAGAGATGGGGAAGACTATATCTGCCCAACTG 1220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 AAGCTTCTTCACACGACAGATGGGCGAGAGACAAAGATATGTCTGTAAATG 338
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-09-764-864-471
; Sequence 471, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OR INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 471
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (894)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (903)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (905)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (914)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-471

Query Match      2.0%; Score 51.4; DB 9; Length 915;
Best Local Similarity 69.3%; Pred. No. 6.1e-05;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 1067 CCTGTATGATTTGCGCGCGCTCTCAACAAGATTATGATTGCTGTGACCGCT 1126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 305 CCATCTACTGATCTGCGCAACCGGACATCACTGCTTCATGATCGGGTGTGCAACT 364
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1127 GTGAAGATGGTTTCATGGCGATTTGTGTGGCATTTTGAG 1167
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 365 GCAATGAGTGTTCATGTGGGACTGATCCGATCATCTGAG 405
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-10-175-523-142
; Sequence 142, Application US/10175523
; Patent No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
```

```

; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Paley, Michael
; APPLICANT: Rajan, Prithi
; TITLE OR INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142
; LENGTH: 2320
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-175-523-142

Query Match      2.0%; Score 51.4; DB 15; Length 2320;
Best Local Similarity 69.3%; Pred. No. 0.00012;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 1067 CCTGTATGATTTGCGCGCGCTCTCAACAAGATTATGATTGCTGTGACCGCT 1126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 CCATCTACTGATCTGCGCAACCGGACATCACTGCTTCATGATCGGGTGTGCAACT 192
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1127 GTGAAGATGGTTTCATGGCGATTTGTGTGGCATTTTGAG 1167
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 GCAATGAGTGTTCATGTGGGACTGATCCGATCATCTGAG 233
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: April 23, 2004, 23:50:37
Job time : 1250.38 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 07:20:53 ; Search time 6433.74 Seconds
(without updates)

12114.300 Million cell updates/sec

Title: US-09-787-016a-1
Perfect score: 2610
Sequence: 1 ctcgtgcccgcgcgcacccac.....accttaagatcatatccctg 2610

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estsba:*
2: em_estin:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_p1n:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2539	97.3	2752	11	BC014615
2	1025.2	33.3	4619	11	AK042474
3	1007.2	38.6	1086	9	AL578222
4	1002.6	38.4	4651	11	BC060249

5	957.2	36.7	1093	9	AL555810
6	910.6	34.9	1033	9	AL580669
7	905.6	34.7	1201	9	AL576399
8	900.2	34.5	1010	13	BX380892
9	868	33.3	1201	9	AL514138
10	863.6	33.1	1201	9	AL553387
11	845.2	32.4	1018	9	AL550675
12	830.2	32.1	1050	13	BX404270
13	830.4	31.8	850	14	CD364822
14	817.8	31.3	1614	11	AK014422
15	813.4	31.2	1119	9	AL550431
16	808.6	31.0	1640	11	BC029110
17	798.6	30.6	898	13	BQ646878
18	792.8	30.4	939	9	AL571596
19	786.6	30.1	891	13	BX350413
20	786.4	30.1	1040	9	AL580668
21	786.2	30.1	925	9	AL546121
22	765.8	29.3	1201	13	BX442517
23	753.4	28.9	781	13	BQ921479
24	740.8	28.4	1061	9	AL574291
25	738.8	28.3	842	13	BQ437271
26	733.2	28.1	866	13	BQ212019
27	722.4	27.7	1201	9	AL514137
28	717	27.5	1045	9	AL550704
29	708.4	27.1	740	13	BQ689908
30	693.8	26.6	779	12	BQ825278
31	673.2	25.8	842	12	BG757160
32	670.6	25.7	839	12	BQ828423
33	668.6	25.6	1097	12	BQ290255
34	668.2	25.6	900	13	BX413695
35	648.6	24.9	760	10	BQ298017
36	645	24.7	740	12	B1913500
37	634.4	24.3	963	13	BQ526593
38	629.6	24.1	772	10	BQ299127
39	626.8	24.0	692	10	BQ393471
40	625	23.9	653	13	BQ639998
41	615.6	23.6	653	10	AM373892
42	615.6	23.6	653	10	AM373829
43	614.4	23.5	923	10	BF308603
44	612.4	23.5	1036	13	BQ881380
45	608.8	23.3	868	10	BF306721

ALIGNMENTS

RESULT 1	BC014615	2752 bp	mRNA	linear	HTC 17-DEC-2003
LOCUS	BC014615				
DEFINITION	Homo sapiens CDNA clone IMAGE:3532298, containing frame-shift errors.				
ACCESSION	BC014615	GI:33869935			
VERSION	BC014615.2				
KEYWORDS	HTC				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Dukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 (bases 1 to 2752) Straussberg, R.L., Feingold, R.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shennan, C.M., Schler, G.D., Altschul, S.F., Zeeberg, B., Buecaw, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, Y., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.R., Brownstein, M.J., Ustun, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loguella, N.A., Peters, G.J., Abramson, R.D., Mulhally, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyik, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,				

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITL
JOURNAL

2 (bases 1 to 2752)
Strauberg, R.
Direct Submission
Submitted (24-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 19, 2003 this sequence version replaced gi:15779097.
Contact: MGC help desk
Email: cgapdb-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
Series: IRML Plate: 11 Row: j Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18375616
This clone has the following problem: frame shifted.
Location/Qualifiers
1. 2752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3532298"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_1ib="NIH MGC_17"
/lab_host="DH10B-R"
/note="vector: pOTB7"

ORIGIN

Query Match 97.3%; Score 2539; DB 11; Length 2752;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2573; Conservative 0; Mismatches 0; Indels 4; Gaps 3;

36 GAAATGGCTGCGAGACCTTAGAGGCTGCGAGGCTTACTCCAGGGAACAGCTCTAGAT 95
1 GAAATGGCTGCGAGACCTTAGAGGCTGCGAGGCTTACTCCAGGGAACAGCTCTAGAT 60
96 AATCTGAGTTGTGAATAATACGAAGCTTGTACTGTGAACAGTGGCTGCAACATGTT 155
61 AATCTGAGTTGTGAATAATACGAAGCTTGTACTGTGAACAGTGGCTGCAACATGTT 120
156 GTTGGAGCTGGCTGTCTGCTTGAACCAAGAGGTTGCTGCGAGGGTTTGGTTGT 215
121 GTTGGAGCTGGCTGTCTGCTTGAACCAAGAGGTTGCTGCGAGGGTTTGGTTGT 180
216 ATTAGATTTCAGGGAAGTGTCCAGCTTCACTGTGAGCAAGTATGAGCAAA 275
181 ATTAGATTTCAGGGAAGTGTCCAGCTTCACTGTGAGCAAGTATGAGCAAA 240
276 AGCCGACCCGAGCAATGAGAGGCACTTAGGCACTCAACCAACAGCAAAAGTTTCAG 335
241 AGCCGACCCGAGCAATGAGAGGCACTTAGGCACTCAACCAACAGCAAAAGTTTCAG 300
336 GAAACATATGGGTTTTCGAAAGCACTATGCGCAAGGAGGCGCAGGGGACCGGCA 395
301 GAAACATATGGGTTTTCGAAAGCACTATGCGCAAGGAGGCGCAGGGGACCGGCA 360

QY 336 GGCTGACCCACTGAGAGCCGCCACCCCAAGAGAGGCTGGGCTGTCCCTGGGGCGAG 455
DB 361 GGCTGACCCACTGAGAGCCGCCACCCCAAGAGAGGCTGGGCTGTCCCTGGGGCGAG 420
QY 456 TGGAGAGGAGCCCAAGCCCACTGAGCGGCTGGAGCAAGTTCTTGAACATTTGCGGGCCG 515
DB 421 TGGAGAGGAGCCCAAGCCCACTGAGCGGCTGGAGCAAGTTCTTGAACATTTGCGGGCCG 480
QY 516 CGGCAAGAGAGCAATGCTGTCTCTCTGAGAGATTCTGTGAGCCCACTCTGCCCCG 575
DB 481 CGGCAAGAGAGCAATGCTGTCTCTCTGAGAGATTCTGTGAGCCCACTCTGCCCCG 540
QY 576 CACACAGCCCGAGAACAGCTCTCGAGAGGCAAGCTGGAAGGCTTCTGAGACCAAGAGCG 635
DB 541 CACACAGCCCGAGAACAGCTCTCGAGAGGCAAGCTGGAAGGCTTCTGAGACCAAGAGCG 600
QY 636 CCCCAGCTCTGCTTCAACAGCTGTGAGAGGAAGAACAGCAGCTCTTGTGAAGGTAAG 695
DB 601 CCCCAGCTCTGCTTCAACAGCTGTGAGAGGAAGAACAGCAGCTCTTGTGAAGGTAAG 660
QY 696 AGGGAGATGACCAAGATGACACCTCGGATAGTGAACAGCGATGACCTTGAAGAGCT 755
DB 661 AGGGAGATGACCAAGATGACACCTCGGATAGTGAACAGCGATGACCTTGAAGAGCT 720
QY 756 TGAAGATGCTCTTCCGACGAGAACCGGGAACAGAGCCCAATTGAGAGGCCCTGAAAGGAT 815
DB 721 TGAAGATGCTCTTCCGACGAGAACCGGGAACAGAGCCCAATTGAGAGGCCCTGAAAGGAT 780
QY 816 CCAGAGTGCCTGCGGGAAGAGCGCGGAGAGAGGCTCCGCGAGACTGTGGGCTCGA 875
DB 781 CCAGAGTGCCTGCGGGAAGAGCGCGGAGAGAGGCTCCGCGAGACTGTGGGCTCGA 840
QY 876 GGCAGATGACACTGTGAGAGGAGCTCTGCGAGTGAAGCAGAGCCCGAGAGCATCAAG 935
DB 841 GGCAGATGACACTGTGAGAGGAGCTCTGCGAGTGAAGCAGAGCCCGAGAGCATCAAG 900
QY 936 GGTGTGTCTCCAGCGCTGGGAAAGATGACAGAGAGTAAAGTTGAGAGGAAGCGGCTCA 995
DB 901 GGTGTGTCTCCAGCGCTGGGAAAGATGACAGAGAGTAAAGTTGAGAGGAAGCGGCTCA 960
QY 996 GGAACATCAAGATGAGAGAGCTTGAAGCTTGGGCGAGCCGAGAGCTGAATGTGAGGTTA 1055
DB 961 GGAACATCAAGATGAGAGAGCTTGAAGCTTGGGCGAGCCGAGAGCTGAATGTGAGGTTA 1020
QY 1056 CGAACCCCAAGCGCTGTATTTGCAATTTGCGGCAAGCTTCAACACAGGTTATGATTTG 1115
DB 1021 CGAACCCCAAGCGCTGTATTTGCAATTTGCGGCAAGCTTCAACACAGGTTATGATTTG 1080
QY 1116 CTGTGACCGCTGGAAGATGCTTTCATGAGCAATTTGTGAGGCTGAGG 1175
DB 1081 CTGTGACCGCTGGAAGATGCTTTCATGAGCAATTTGTGAGGCTGAGG 1140
QY 1176 GAGGCTTTTGAAGAAAGGATGAGGAAAGCTATATCTGCCAACTGACCAATTTGCAAGT 1235
DB 1141 GAGGCTTTTGAAGAAAGGATGAGGAAAGCTATATCTGCCAACTGACCAATTTGCAAGT 1200
QY 1236 GCAGATGAGACTCATTCAGAAACGCGAGATACAGAGAACTTAAATGAGACTTGAGA 1295
DB 1201 GCAGATGAGACTCATTCAGAAACGCGAGATACAGAGAACTTAAATGAGACTTGAGA 1260
QY 1296 TGTGTATGCAACGATTTGTATGAAATAGGAAACAAATAGAGCAGAGCTTAGAGAGCA 1355
DB 1261 TGTGTATGCAACGATTTGTATGAAATAGGAAACAAATAGAGCAGAGCTTAGAGAGCA 1320
QY 1356 AGGAGTAAAGGGTATGAGAAAGCTGCAATTCAGAGTGGCAGAGAACTCAAGAT 1415
DB 1321 AGGAGTAAAGGGTATGAGAAAGCTGCAATTCAGAGTGGCAGAGAACTCAAGAT 1380
QY 1416 CTTCAGGCTGTGATGAGGCGGCTGTGCTCAAAATGATTTGCGCGGGGTCTGTCA 1475
DB 1381 CTTCAGGCTGTGATGAGGCGGCTGTGCTCAAAATGATTTGCGCGGGGTCTGTCA 1440

1476 CGTGGCGAGCCGAGCTGGTGTACTGCAATGATCTGTATCTCTCAAAACAGCCGCGAC 1535
 1441 CGTGGCGAGCCGAGCTGGTGTACTGCAATGATCTGTATCTCTCAAAACAGCCGCGAC 1500
 1536 GACAAATGAAGTTTCTTAAGCTCAGGTAAGAAACAGAAAGCCAAAGCCTAAAGAAAGATGA 1595
 1501 GACAAATGAAGTTTCTTAAGCTCAGGTAAGAAACAGAAAGCCAAAGCCTAAAGAAAGATGA 1560
 1596 GATGAAGCCAGAAAGCCAGCTCTCCGAAATGCGGTCTCAGGAGGATTAATTAATCTC 1655
 1561 GATGAAGCCAGAAAGCCAGCTCTCCGAAATGCGGTCTCAGGAGGATTAATTAATCTC 1620
 1656 TTCTGTGCAAGAGAGCAGCTCCAG-AAAAAAAAAGACCAAGTGAAGAAAGCAGTGG 1714
 1621 TTCTGTGCAAGAGAGCAGCTCCAGAAAAAAAAAAAAAAAAAGACCAAGTGAAGAAAGCAGTGG 1680
 1715 TGGTCCCTGCGCGAGTGAAGCACTGGGAAAGAAAGCCTTGTAGAGACAGCAGCCGT 1774
 1681 TGGTCCCTGCGCGAGTGAAGCACTGGGAAAGAAAGCCTTGTAGAGACAGCAGCCGT 1740
 1775 CGTGGCGAGCCGAGCTCAGATTCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAG 1834
 1741 CGTGGCGAGCCGAGCTCAGATTCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAG 1800
 1835 CGTCACTGTGTATAATGTATGTATACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1894
 1801 CGTCACTGTGTATAATGTATGTATACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1860
 1895 CTTTCTGTGATGAGCCATCCCTGGGCTGTGCAAGCACTGGGAGTGTGAGCTTGTGTAG 1954
 1861 CTTTCTGTGATGAGCCATCCCTGGGCTGTGCAAGCACTGGGAGTGTGAGCTTGTGTAG 1920
 1955 CTGATCAAGAGACCGGCTGCAACATCAGCGGAGAGCAGAGCCCATATGTCAGAGATCCTC 2014
 1921 CTGATCAAGAGACCGGCTGCAACATCAGCGGAGAGCAGAGCCCATATGTCAGAGATCCTC 1980
 2015 CTGCTGCCCTGTGTCTCATCTTCAAGCACTTCTGTCACTGTCTTCAAAAGCTG 2074
 1981 CTGCTGCCCTGTGTCTCATCTTCAAGCACTTCTGTCACTGTCTTCAAAAGCTG 2040
 2075 TAAACTCACTGTGTGAAGCTTCACTTATGATGATGATGATGATGATGATGATGATGATGATG 2134
 2041 TAAACTCACTGTGTGAAGCTTCACTTATGATGATGATGATGATGATGATGATGATGATGATG 2100
 2135 CAGGCTCTGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2194
 2101 CAGGCTCTGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2160
 2195 CGGCAAGCCAGATCCACCCCTGTGTGCAATGATGATGATGATGATGATGATGATGATGATGATG 2254
 2161 CGGCAAGCCAGATCCACCCCTGTGTGCAATGATGATGATGATGATGATGATGATGATGATGATG 2220
 2255 CTTCACTTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2314
 2221 CTTCACTTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
 2315 GTGGGAATCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2374
 2281 GTGGGAATCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
 2375 GTGGGAATCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2434
 2341 GTGGGAATCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2398
 2435 AACCACTTGTGAAGCTTGAAGGCAACCTCTGTCTCATAGGCCCCCAGGTAAGAGT 2493
 2399 AACCACTTGTGAAGCTTGAAGGCAACCTCTGTCTCATAGGCCCCCAGGTAAGAGT 2458
 2494 TCACACTTGAAGCACTGTCTGTGACCCGAGAGCCGCTGCTTGAAGCTTGTGATCTTCAAG 2553
 2459 TCACACTTGAAGCACTGTCTGTGACCCGAGAGCCGCTGCTTGAAGCTTGTGATCTTCAAG 2518
 2554 TGAAGGCTTTTCTTGTGCTGTGTCTTGAATGTTTGAAGCTTGAAGTCAATATCTG 2610

Db 2519 TGAAGGCTTTTCTTGTGCTGTGTCTTGAATGTTTGAAGCTTGAAGTCAATATCTG 2575
 RESULT 2
 AK042474
 LOCUS
 DEFINITION
 Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone: A630095A07 product: death inducer-oblierator-1, full insert sequence.
 AK042474
 VERSION
 AK042474.1 GI:26335138
 HTG, CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS
 Carrinci, P. and Hayashizaki, Y.
 TITLE
 High-efficiency full-length cDNA cloning
 JOURNAL
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE
 99279253
 PUBMED
 10349636
 REFERENCE
 AUTHORS
 Carrinci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE
 20499374
 PUBMED
 11042159
 REFERENCE
 AUTHORS
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carrinci, P., Komuro, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL
 Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE
 20530913
 PUBMED
 11076861
 REFERENCE
 AUTHORS
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 TITLE
 Functional annotation of a full-length mouse cDNA collection
 JOURNAL
 Nature 409, 685-690 (2001)
 REFERENCE
 AUTHORS
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL
 Nature 420, 563-573 (2002)
 REFERENCE
 AUTHORS
 Adachi, U., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carrinci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasuwa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komuro, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, K., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

COMMENT

Kanagawa 230-0045, Japan (E-mail: genome-res@isc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL: <http://genome.gsc.riken.go.jp/>
 URL: <http://fantom.gsc.riken.go.jp/>
 Location/Qualifier

FEATURES

source

CDS

1..4619
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:A630095A07"
 /db_xref="MG1:2407519"
 /db_xref="taxon:10090"
 /clone="A630095A07"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"
 299..2143
 /note="unnamed protein product; death
 inducer-obitator-1 (MGD|MG1:1344352, GI:A238332,
 evidence: BLASTN, 99%, match=2599)
 putative"
 /codon_start=1
 /protein_id="BAC31270.1"
 /db_xref="GI:26335139"
 /translation="MDDKGLNBRAPKIKPTSKFRRTMGPRRTIARSGADTE
 VPSRQPOQHLSLRSGRQKRTREVEPTTTRRKKKVPVLEPSRSTV
 TDVEATSESVSSIRSGPDSLGKHPASSEBAKAGEBEDTSDSDGLTKE
 LQNLRRKROBVERSLSQNLKRRRBDASVQJGSADDPICKQBEA
 SCGPVSEETDIEQLBGKATQNTBEPREKPRCEYDPAALYCICRPHNN
 RMIICDRBEWFGDVGISEARGLKRNGBDYICPCTLIQVDEFTSGATDAP
 SCRSVAGDTGCTSIGTIBKSGEDQIKGRLEKRNAGSKKKLKIPOPVEAPAP
 KCIPOCSVADPDPVYCSNDCLIRAAATMRPLSGKOKTKPKKTKPKSLP
 KCSVVGITKISSVHKLASKEKENPVKATMLASRSTSGKRAACESSTTSMASDHNYN
 AVKPEKPTALSPILLSKCTHPKAGPSPSHLGGCLGSRTRVLVLVLAASS
 SIIPARSRYDASGPQVFLPFLMSLSGMLFKSCVGLMLRAISYFSRPM"

ORIGIN

Query Match 39.3%; Score 1025.2; DB 11; Length 4619;
 Best Local Similarity 71.9%; Pred. No. 1,4e-243;
 Matches 1495; Conservative 0; Mismatches 533; Indels 50; Gaps 10;
 21 TCCGCGGCGCTTGGGGAATGCTGCGAGACCTTAGAGGCTCGGAGCTTACCAACG 80
 68 TCCGTGTAGCTCTGGAATGCTGCGAGATCCCGCGCGCGGAGCTTTTAAGAGG 127
 81 GAACAGCTCTAGATATCTGAGTTTGAATAATACGAAGCTGTACTCGTAACAGTG 140
 128 CAGTCCCACTGCTCCCTTGTGTTCGAAGCTCCGAATCTTCTCATTTGATGAACGTG 187
 141 GCTGACAAAGTGTGTGTGAGCTCTGCTGTCTTGAGACCAAGGTTTGTCTGCC 200
 188 ACTGACAAAGTGTGTGTGAG- GCTTGTGCTGTCTTAC-----TGACC 233
 201 AGGGTTTGTGTGTATTAGATT- CAGGGAAGATGTCAAGCTTTCAGTGTGAG 259
 234 CAGGTATTAATTTTATGAGATTTCACGCAAGATTTTCAAGCTTTCAGTTTGGGA 293
 260 CAGGTATGACCAAGGCGAGCCGAGCAATGAGAGCACTTAAGCCATCAAAACCA 319
 294 CAGGTATGATATTAAGGCACTTGAGCAATGAGAGCACTTAAGCCATCAAAACCA 353
 320 CAGCAAAAGTTTCAAGAAACATGTGGGTTTTCGAAGACCACTATTCGCAACGAGAG 379

DB 354 CAGGTATGAGATTTCAAGAAACCTTGGGTTTTCGAAGACCACTATTCGCAACGAGAG 413
 380 GCGCAGAGGAGCGCGAGGCTGACCTCACTGAGCCGACACCCCAACAGCAGACTGGGCC 439
 414 GTGCAAGAGACACGAGAGGTGAGACCCCACTGAGAGCAAC-----ACAGACATTAAC 467
 440 TGTCCCTGTGGGCGAGTGGAGGAGCCCAAGGCACTGAGGCGTGGAGCAGTTCCTGA 499
 468 TCTCCCTGTGGCGAGTGGAGCGCAACCAAGTACTGAGAGGTTGAAGATTTCTTA 527
 500 CCAATTCGCGGCGCGCGCAGAGAGAGATGCTGTCTCCCTGAGAGATTCTGTGAGC 559
 528 CCAAGTGTGGCGCGCGAGGAAAGAAATGTGCGGTGTCTCGAGAGATTTCAGTGAGC 587
 560 CCAAGTGTGGCGCGCGCAAGAGCGCCGAGACAGCTTCGAGGCGAGCTGGAAAGGCTT 619
 588 CCAATCTTTCACAGTCACTGATGTGAGACAGCTTCCGAGGAGAGCTTGAAGCAGTT 647
 620 CTGAGACCAAGAGCGGCGCGCGCTGTCTTCCACAGCTGTGAAGAAACAGCGCTCTT 679
 648 CTGAGATCAGAAAGTGCCTGTATCTGACTCTTA--GGAAAGAACATCTGTGCTCTT 704
 680 CTGAAAAGGTGAAGAGGAGATGACCAAGATGACACTCCGATAGTGAACAGGATGCC 739
 705 CTGAAAAGCAAAAGAGATGAGAGAGAAAGAAACACTCTGACATGACAGATGAGCC 764
 740 TGAAGCTTGAAGAGCTTCAAGATGCTTTCGAGAGAGCGGAAACAGAGCCCACTGAGA 799
 765 TTAAGTTGAAGAACTTCAAGAACCGCTTCGAGAGAGCGAGAGCAAGAACTGTGAGA 824
 800 GCGCCCTGAAGAGATTCAGATGCTGCGGAGAAAGCGCGGAGAGAGGCTCCGCG 859
 825 GGTCTCTGAGAGCAATCAGAAATGCTTGAAGAAAGCGCAAGAGAAAGATTCTGCG 884
 860 AGACTGTGAGCTCCGAGCGCAAGTGAACATGTGAGAGGCGCTCTGCCAGTAAAGAGAGC 919
 885 AATCTGAGATGCTCAAAATAGGAGTGCAGAGAGAGAGAGAGAGCTCTCTTAAGAGAGAGC 944
 920 CCGAGAACATCAAGGGGTTGTGTCCGAGCTGAGAAAGTGAACAGAGAGATTAATTGG 979
 945 CTGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGAGCAGATGACATTAAGAAATCAGTTGG 1004
 980 AGGGAAGGCGCTCGAGACATCAAGATGAGAGGCTGAGAGCTTGGGCGAGCGAGAGC 1039
 1005 AAGGAGAGGCGACTCGAGAAATACAGAGAGAAACCCAGGAGAGCGGAGAACCAAGC 1064
 1040 CTGAATGTAGAGGTTACAGACCCCAAGCGCTGTATTGCAATTTGCGCAGACTTCAACA 1099
 1065 CTGAGTGTAGGTTTACAGACCCCAAGCGCTGTACTGCAATCTGCGCAGACTTCAACA 1124
 1125 ACAGGTTTATGATCTCTGTGATGAGTGTGAGAGAGTGTCTCAATGTGATCTGTGTG 1184
 1160 TTTCTGAGGCTCGAGGAGGCTTTTGAAGAGAAATGGGAAAGCTATATCTGCCAAACT 1219
 1185 TTTCTGAGGCGCGAGGCGGCTCTGGAAGAGAAAGGAAAGCTATATCTGCCAAACT 1244
 1220 GCAACATTTTTCAGAGTGAAGTGAAGTCAATTCAGAAACGAGAGTCAAGAGAACTA 1279
 1245 GCAACATTTTTCAGAGTGAAGTGAAGTCAATTCAGAAACGAGAGTCAAGAGAACTA 1304
 1280 AATGAGACTGTGAGATGCTGATGAGCAAGATTTTCAAGTATAGGAACAATAGAGCA 1339
 1305 GGTGAGATCTGTGAGTGTGATGAGCAAGATTTTCAAGTATAGGAACAATAGAGCA 1364
 1340 AGTCTAGCAAGGCAAGGATTAAGGTTAAGTTGAGAAAGCTGCAAAATTCAGAGTGA 1399
 1365 AGTCTAGCAAGGCAAGGATTAAGGTTAAGTTGAGAAAGCTGCAAAATTCAGAGTGA 1424
 1400 AGAAGAACTCAAGATCTTTCAGAGCTGTGATAGAGCGCTGTGCTCAAAATGATTTG 1459

Db	1425	AGAAAAAACTGAAGTATTCAGAGCTGTGTAAGAGGCTCTGGTGTCTCTAAATGATTC	1484
Qy	1460	GCCCCGGGTGTGTCACTGTGGCGACAGCCCGACTGGGTACTGCAGTAAATGATGTATCC	1519
Db	1485	GCCCTGGGTGTTCAGTGTATGACACAGCTGTGACTGTGTATTTGGCAGTAAATGATGTATTC	1544
Qy	1520	TCAAAACGCGCGACGCGCAATGAAAGTTCTTAAGCTCAGGTAAAGAAACAGAAAGCAAAAGC	1579
Db	1545	TCAAAACAGCGCAGCTACCAATGAAATTTCTAAGTTACAGTTAAAGAAACAAAACAAAAC	1604
Qy	1580	CTAAAGAAAAATGAAGATGAAGACCAAGAAAGCCAGACTTCCGAAATATGGGTCTCAGG	1639
Db	1605	CCAGGAAAGAGCTCAAGACGAAAGCCAGAAAGTTTCAAGCTTTCGAAATATCAGTGTTCAGG	1664
Qy	1640	CAGGTATTTAAATCTCTTCTGTGTCAAGAGAACCACTCTCAGAAAAAAAAGACCAACAG	1699
Db	1665	TGGGGATTAAATCTCTTCTGTGTCAAGAGAACCACTCTCAGAAAAAAAAGACCAACAG	1724
Qy	1700	TGAAGAAAGGCACTGTGTGTCTCTGTGGCGAGTGAAGCACTGGGGAAGAGACAGCTTGTG	1759
Db	1725	TGAAGAA---AGTATGTCTGGCTTCCAGAGTGAAGCTTCTGGGGAAGAGGACAGCTGTG	1781
Qy	1760	AGACGACGACGCGCGTGTGGGCGAGAGCATCAATTAATGAGTAAGCCAGAAAAA	1819
Db	1782	AGACGACGACCACTCTGTGGGAGATGACCAACTAATAAGTCTGTGAAGCCAGAGAAC	1841
Qy	1820	CTG-----CTGCTCCCTGCGCGTCACTGTGTATTAATGATATGATCACTAGGGG	1870
Db	1842	CAGAGAAAGCCACTGTCACTCTGTGCCAACCCTATTGATGAATGTATGATCAACCCAAAG	1901
Qy	1871	TTGGCTCTCTGAGACCCCTCCGGTCTTTCTGGAATAGCAATCCCTGGGCTGTCCAGAGC	1930
Db	1902	CTGGCTTCCAGAGCCCTCCCATCATCT-----GGTGTGCTGCTGGGGCTGTCTAGAAC	1956
Qy	1931	TGGAGTTGACAGCTTGTGTTAAGCTATCAACAGACATCGGCTGCACATACAGGGGAG	1980
Db	1957	CAG-----AGTCTGGGGTGTCTGTGTGTGATGATGAGCCAGAC---TCACTGCCAGC	2007
Qy	1991	CAGAGCCCATGTCCAGAGATGAGCTCGCTGTGTGCTCTGTGTGCATATCCCTAGTGTGAGACT	2050
Db	2008	CAGAAAGCAGATACAGAGATGCTCTGTGACCCGAGGTGTCTGTGCTTACCTGTGTGAGCTT	2067
Qy	2051	TCTGTCACTGTTTTCAAAGCTGTAAACCTCACTGT 2088	
Db	2068	CTCTGGTGTCTCTTAAGAGCTGTGTAGGCTCATGT 2105	
RESULT 3			
AL578222/c			
LOCUS	AL578222	1086 bp	mRNA linear EST 01-JUN-2003
DEFINITION	AL578222 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens		
ACCESSION	AL578222		
VERSION	AL578222.2		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Buhalaxota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Li, M.B., Gruber, C., Jessup, J., and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
On Feb 16, 2001 this sequence version replaced gi:12942093.			

```

FEATURES
source
    Location/Qualifiers
    1..1086

```

ORIGIN

Query Match	38.64;	Score 1007.2;	DB 9;	Length 1086;
Best Local Similarity	97.54;	Pred. No. 2.8e-239;		
Matches 1047;	Conservative 5;	Mismatches 18;	Indels 4;	Gaps 3;

QY	1047	TGAGGGTTAGCAACCCCAACCCCGTATGTCATTTGCGCAGGCTCAACAACAGGTT	1106
Db	1072	TGTGAGGTTACGACCCCAACCCCGTAT --CGTTGCGCAGGCTCAACAACA-GTT	1016
QY	1107	TATGATTTCTGTGTGACCGCTGTGTAGAAATGTTTTATGCGCATTTGTGGCATTTTCGA	1166
Db	1015	TATGATTTGCTGTGACCGCTGTGTAGAAATGTTTTATGCGCATTTGTGTGGCATTTTCGA	956
QY	1167	GGCTCGAAGGGAAGGCTTTTGGAAAAGAAATGGGAAAGATATATCTGCGCCAACTGCAACAT	1226
Db	955	GGCTCGAAGGGAAGGCTTTTGGAAAAGAAATGGGAAAGATATATCTGCGCCAACTGCAACAT	896
QY	1227	TCTGCAAGTGCAGGATGAGACTCATTTCAAGAAACGGCAGATCAAGCAGAAAGCTAAATGAG	1286
Db	895	TCTGCAAGTGCAGGATGAGACTCATTTCAAGAAACGGCAGATCAAGCAGAAAGCTAAATGAG	836
QY	1287	ACCTGAGATGCTGATGTGACCGAATGTACAGATATAGAAACAATAGACGAAGTCTAG	1346
Db	835	ACCTGAGATGCTGATGTGACCGAATGTACAGATATAGAAACAATAGAGCAGAAGTCTAG	776
QY	1347	CGAAGACCAAGGGATTAAGGGGTAGAAATTTGAGAAAAGCTGCATAATCCAAATGTGCCAAGAA	1406
Db	775	CGAAGACCAAGGGATTAAGGGGTAGAAATTTGAGAAAAGCTGCATAATCCAAATGTGCCAAGAA	716
QY	1407	ACTCAAGATCTTCACGCTGTGTGATPAGAGGGCGCTGTCCTCAAAATGTATTTGCCCCGG	1466
Db	715	ACTCAAGATCTTCACGCTGTGTGATPAGAGGGCGCTGTCCTCAAAATGTATTTGCCCCGG	656
QY	1467	GTCCTGTCAAGTGGCGCAGCCGCACTCGGTGTACTGCAATATGACTGTATCTTCAACA	1526
Db	655	GTCCTGTCAAGTGGCGCAGCCGCACTCGGTGTACTGCAATATGACTGTATCTTCAACA	596
QY	1527	CGCCGAGCGACAATGAAATTTCTTAAGCTCAAGTAAAGAACAGAGCCAAAGCCTTAAGA	1586
Db	595	CGCCGAGCGACAATGAAATTTCTTAAGCTCAAGTAAAGAACAGAGCCAAAGCCTTAAGA	536
QY	1587	AAAGATGAAGTAAAGCAGAGAAAGCCCACTTTCCGAAATGCGGTCTCAGGACAGAT	1646
Db	535	AAAGATGAAGTAAAGCAGAGAAAGCCCACTTTCCGAAATGCGGTCTCAGGACAGAT	476
QY	1647	TAAAAATCTCTTCTGTGTCAACAAGACAAGGCTCCAGAAAAAAAAGAGACAAGTGAAGAA	1706
Db	475	TAAAAATCTCTTCTGTGTCAACAAGACAAGGCTCCAGAAAAAAAAGAGACAAGTGAAGAA	416
QY	1707	GCGAGTGTGTGTCCTTCGCGGAGTGAAGACTTCGGGAAGAAAGCACTTGTGAGCAG	1766
Db	415	GCGAGTGTGTGTCCTTCGCGGAGTGAAGACTTCGGGAAGAAAGCACTTGTGAGCAG	356
QY	1767	CAGCGCGTGTGTGGCTCAGGCAATTAACAATGCAATTAAGCACAABAAAAGACTGTCG	1826
Db	355	CAGCGCGTGTGTGGCTCAGGCAATTAACAATGCAATTAAGCACAABAAAAGACTGTCG	296

OY	774	GAAGCGGGAAACAGAGCCCACTGAGAGGCCCTCTGAAAGGGATTCAGAGTGGCCCTGGCGGAA	833
Db	778	AAAGCGAGAGAGAAACCTGTGGAGAGGATCTCTGAGAGGCACTGCAAAATCCCTGAGGAA	837
OY	834	GAAGCGCGGGAGAGAGGGTCCCGCGAGACTGTGGGCTCCGAGGCCAGTGTGACACTGTGGAA	893
Db	838	GAAGCGGAGAGAGAAAGATTCTGGCGAAACTGTGGAGTGTCCAAATATGGCAAGTGCAGAGCA	897
OY	894	GGGCGTCTGCCAGTAAAGCAGAGAGCCCGAGAACATCAGAGGGGTGTGTCTCCAGGCTGG	953
Db	898	GGACAGACCTCTCTGTAAAGCAGAGGCTCGAGGCTAGTACAGGAGCAAGGTGTCCAGTCAAGA	957
OY	954	GAAAGATGACAGAGAGATAGTTGTAAGGAGAAAGCGCGCTCAGAGCATCAAGAATGAGAGA	1013
Db	958	GACAGATACATAGAAATTCAGTTGTAAGGAGAGGCCATCTAGGAAATATCAGAGGAAAA	1017
OY	1014	GCCTTGAGACTTGGGCGCAGCCGAAGCCCTGAATTGTGAGGGTTACGACCCCAAGCCCTGTGA	1073
Db	1018	CCCGAGGGAGACGGGGCAAAACAAAGCCCTGAGTGTGAGGTTTACGACCCCAATGCGCTGTGA	1077
OY	1074	TTGCAATTTGCGCGCCAGCCTTCACAAACAAGGTTATATGTTTGCTGTGACCGCTGTGAGAA	1133
Db	1078	CTGCAATCTGCGCGCAGCCTTCACAAACAAGGTTATATGATCTGTGTATCGGTGTGAGAA	1137
OY	1134	ATGGTTTCATGGCGGATTTGTGTGGGCAATTTCTGAGGCTGAGGAGAGGCTTTTGGAAAGAA	1193
Db	1138	GTGGTTCCATGGTATCTGTGTGGGATTTCTGAGGCGCGGAGGCGGCTCTGTGAAAGGAA	1197
OY	1194	TGGGGAGACTATATCTGCGCCCAACCTGACCAATTTGTGCAATGTGACAGATCTGACATTC	1253
Db	1198	CGGGAGAACATCAATCTGCGCCCAATTTGACCAATTTGTGCAAGTGTGACAGATCTGACATTC	1257
OY	1254	AGAAACGGCAGATCGACAGAGGCTTAAATGTGAGACTTGAAGTGTCTGATGTGCACCGATTG	1313
Db	1258	TAGCGCCACCGATGTAGCAGAGCTCTGGGTGCGAGATCTGTGGGTGTATGTGCACAGACTG	1317
OY	1314	TACAAATATGAGAACATATGAGCAGAGATCTAGCGAGAACCAAGGGAATTAAGGGTTAGAT	1373
Db	1318	CACAAAGCATATGAGGACATGTGAGCAGAGATCTCGAGAGAACCAAGGGAATTAAGGGTTAGAT	1377
OY	1374	TGAGAAAGCTCCAAATCCAGTGTGGCAAGAGAACTCAAGATCTTCCAGCTGTGATGTAGA	1433
Db	1378	TGAGAAAGCAGCAAAACCCGAGGGGAGAGAAATCTCAAGATTTCCAGCTGTGTGTAGA	1437
OY	1434	GAGCGCTGTGCTCTCAAAATGTATTTGGCCCGGGGTCTGTCAAGTGTGCAGCCCGACTC	1493
Db	1438	GGCTCTGTGTCTCTTAATATGCAATTTGGCCCTGGGGTCTTCAAGTGTGACAGCCCTGACTC	1497
OY	1494	GGTGTATCTGCAGTATATGATCTGTATCTTCAAAACAGCCCGCAGCAGCAATGAAGTTCTTAG	1553
Db	1498	TGTGTATTTGCAGTATATGTACTGTATCTTCAAAACAGCCAGCAGCTACCATGAGATTTCTTAG	1557
OY	1554	CTCAGGTAAAGAAC-AGAAAGCCAAAGCTTAAAGAAAGATGAAAGATGAAAGCCAGAGAAC	1612
Db	1558	TTCAAGTAAAGAACCAAAAACCAAAACCAAGAGAAAGTCAAGCAGAGCCAGAAAGT	1617
OY	1613	CCAGTCTTCCGAATGCGGTGTCTAGCGAGGATTTAAATCTCTTCTGTGCAAAAGAAC	1672
Db	1618	TCAGTCTTCCAAATATGCAAGTGTTCAGGTGGGGATTTAAATCTCTTCTGTGCAAAAGAAC	1677
OY	1673	CAGCTCCAGAAAAAAGAGAGCAACAGTGAAGAAAGCAGTGGTGTCCCTGGCGCGAGATG	1732
Db	1678	TAGCTTCAGAGAAAGAGAGAAACCAAGTGAAGAA--AGTATGTGTGCTTTCAGAGATG	1734
OY	1733	AAGCACTGGGAGAGAGCAAGCTGTGTGAGAGCAGACGCGCTGTGGCGAGAGATCA	1792
Db	1735	AGACTTCTGGGAAAGAGGACAGCTGTGAGAGCAGACCAATCTGGCGCAAGTGAACACA	1794
OY	1793	ATTCAATGCAGTAAAGCCAGAAAAGACTG-----CTGCTCCCTGGCGCTCACTGT	1843
Db	1795	ACTCAATATGCTGTGAAGCCAGAGAAAGCCAGAGAACCCCACTGCAATCTTGCGCCACCTAT	1855

Qy	1844	TGATATAATCTATGATATACCTTAGAGGGGTGGCCCTCCGAGACCCCTCCGGTCTTTCTGGA	1903
Dy	1855	TGATATAATCTATGATATACCTTAGAGGGGTGGCCCTCCGAGACCCCTCCGGTCTTTCTGGA	1909
Qy	1904	TAGCCATCCCTCGGGCCCTGTCCAGACTGGAGAGTTGACGCTTTGTGTTAAGTGATACAA	1963
Dy	1910	GGGAGGCTGCCTGGGGCTGTGCTTAGACACAG-----AGTCCCTGGGGTGTTCGGAGCTGATATA	1964
Qy	1964	GACACCGGCTGCGACATCAAGGAGGAGACAGCCCATATGCTCAGAGATCCCTCGTCTGCC	2023
Dy	1965	GTAGCCAGCAGC-----TCACTGCGACGCCAAGACATACCAAGATGCCCTCTGGACCCCA	2020
Qy	2024	TGATCATCCCTATGATGTACAGACTTCTCTGTCACTGTTTCCAAAGCTGTAAACCTCA	2083
Dy	2021	GGATTTCTCTACTAGCTCTGTGAGAGCTCTCTGGGTGGTGTCTTAAGAGCTGTGTAGGCTT	2080
Qy	2084	CTGTGT 2088	
Dy	2081	CATGT 2085	

RESULT 5

LOCUS	AL555810	1093 bp	mRNA	linear	EST 31-MAY-2003
DEFINITION	AL555810 Homo sapiens HELA cells COT 25-NORMALIZED				
ACCESSION	AL555810				
VERSION	AL555810.2				
KEYWORDS	EST.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				

REFERENCE

REFERENCE 1 (bases 1 to 1093)
AUTHORS I.I. M.B., Gruber, C., Jessse, J., and Polayer, D.
TITLE Pull-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12897895.

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
<http://www.genoscope.cns.fr>
csl-bim/cluster.cgi?seq=CS0DK002C807QPl&cluster=9799.r. Contact :
Feng Liang Email: fliang@lifetech.com URL: <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DK002C807QPl.

Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="takon:3606"
/clone="CS0DK002J13"
/cell_type="HELA CELLS COT 25-NORMALIZED"

```

ORIGIN

Query Match	36.7%	Score	957.2	DB	9	Length	1093
Best Local Similarity	97.8%	Pred. No.	7,8e-27				
Matches 1011, Conservative	3	Mismatches	11	Indels	9	Gaps	4

41 GGCTGCGAGACCTTAGAGGCTCTGCG-----GAGCTTACTCCAGGGGAA CAGGCTCTGAGT 95
OY
57 GGATGCCAGACCTTAGAGGCTCTGCGGCTCGGACTTACTCCAGGGGAA CAGGCTCTTAAT 116
Db

QY 96 AATCTGAGTTGTTAAATACGAAGCCCTGTTACTGCTGAACAGTGGCTGACACAGCTTT 155
 DB 117 AATCTGAGTTGTTGAAAAATACGAAGCCCTGTTACTGCTGAACAGTGGCTGACACAGCTTT 176
 QY 156 GTTGTGAGCCCTGGCTGCTGCTGAGACCCAGAGGTTTGTCTGCTGACAGGGTTTGTGTTGT 215
 DB 177 GTTGTGAGCCCTGGCTGCTGCTGAGACCCAGAGGTTTGTCTGCTGACAGGGTTTGTGTTGT 236
 QY 216 ATTTAGATTTTCAAGGAAAAAGTGTCCAAAGCTTTCAAGTGTGAGAGAGGATATGACGACAA 275
 DB 237 ATTTAGATTTTCAAGGAAAAAGTGTCCAAAGCTTTCAAGTGTGAGAGAGGATATGACGACAA 296
 QY 276 AGCGGACCCGAGCAATGAGAGGCACTTAAGGCTTCAAACTCCACAGCAAAAGATTGAG 335
 DB 297 AGCGGACCCGAGCAATGAGAGGCACTTAAGGCTTCAAACTCCACAGCAAAAGATTGAG 356
 QY 336 GAAAAATGAGGGTTTTCGAGAGGACCACTATGCGCAAGAGGAGAGGAGGAGGAGGAGGAGG 395
 DB 357 GAAAAATGAGGGTTTTCGAGAGGACCACTATGCGCAAGAGGAGAGGAGGAGGAGGAGGAGG 416
 QY 396 GAGCTGACCCACTGAGAGCCGACCCCAACAGAGAGCTGAGGCTGTCTCTGAGGAGGAG 455
 DB 417 GAGCTGACCCACTGAGAGCCGACCCCAACAGAGAGCTGAGGCTGTCTCTGAGGAGGAG 476
 QY 456 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 515
 DB 477 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 536
 QY 516 GGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 575
 DB 537 GGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 596
 QY 576 CACAGACCCGAGAGAGGCTCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 635
 DB 597 CACAGACCCGAGAGAGGCTCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 656
 QY 636 CCCCCAGTCTGTTTCAACAGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 695
 DB 657 CCCCCAGTCTGTTTCAACAGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 716
 QY 696 AGGAGATGACCAAGATGACCACTCGATATGATGACAGGATGAGCTGATGAGAGAGCT 755
 DB 717 AGGAGATGACCAAGATGACCACTCGATATGATGACAGGATGAGCTGATGAGAGAGCT 776
 QY 756 TCAGATGACCTTTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 815
 DB 777 TCAGATGACCTTTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 834
 QY 816 CCAGAGTGGCTTGGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 875
 DB 835 CCAGAGTGGCTTGGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 894
 QY 876 GGGCAGTGAACCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 934
 DB 895 GGGCAGTGAACCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 954
 QY 935 GGGTGTGTCTCCAGGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 994
 DB 955 GGGTGTGTCTCCAGGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 1014
 QY 995 AGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 1054
 DB 1015 AGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 1073
 QY 1055 AGGAGCCCAAGGCGC 1068
 DB 1074 AGGAGCCCAAGGCGC 1087

RESULT 6
 AL580669 1033 bp mrvA linear EST 01-JUN-2003
 LOCUS AL580669 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
 DEFINITION

Homo sapiens cDNA clone CS0DJ007YD06 5-PRIME, mRNA sequence.
 AL580669
 VERSION AL580669.2 GI:31318936
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCES Li, W.B., Gruber, C., Jeessee, J. and Polayes, D. (bases 1 to 1033)
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12946913.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 9799.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DJ007YD0603P1&cluster=9799.r. Contact :
 Feng Huang Email : fliang@life.com URL :
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DJ007YD0603P1.

FEATURES

source
 Location/Qualifiers
 1..1033
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ007YD06"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_line="JURKAT"
 /clone_11b="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 34.9%; Score 910.6; DB 9; Length 1033;
 Best Local Similarity 97.0%; Pred. No. 3,1e-215;
 Matches 934; Conservative 14; Mismatches 9; Indels 6; Gaps 2;
 QY 33 GGGGAAATGCTGCGAGACCTTGAAGGCTTGGAG-----CTTACTCAAGGGAACAGC 87
 DB 67 GGGGAAATGCTGCGAGACCTTGAAGGCTTGGAG-----CTTACTCAAGGGAACAGC 126
 QY 88 CTCTAGATTAATCTGAGTGTGTAATAATCGAAGCCTGTACTCGTGAACAGTGGCTGACA 147
 DB 127 CTCTAGATTAATCTGAGTGTGTAATAATCGAAGCCTGTACTCGTGAACAGTGGCTGACA 186
 QY 148 ACAAGTGTGTTGAGAGCTGCTGTCTGTGAGACCCAGAGGTTTCTGTGCCAGGGTTT 207
 DB 187 ACAAGTGTGTTGAGAGCTGCTGTCTGTGAGACCCAGAGGTTTCTGTGCCAGGGTTT 246
 QY 208 TTGTTGTATTTAGATTTCAAGGAAAAAGTGTCCAAAGCTTTCAAGTGTGAGAGAGGATAG 267
 DB 247 TTGTTGTATTTAGATTTCAAGGAAAAAGTGTCCAAAGCTTTCAAGTGTGAGAGAGGATAG 306
 QY 268 GACGACAAAGGAGACCCAGCAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 327
 DB 307 GACGACAAAGGAGACCCAGCAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 366
 QY 328 GAGTTCAAGAAAAATGAGGTTTTCAGAGGACCACTATGCCAAGGAGAGAGGAGGAGGAGG 387
 DB 367 GAGTTCAAGAAAAATGAGGTTTTCAGAGGACCACTATGCCAAGGAGAGAGGAGGAGGAGG 426
 QY 388 GACGCGAGGCTGACCCACTGAGAGCCGCCACCCCAACAGACAGAGGAGGAGGAGGAGG 447
 DB 427 GACGCGAGGCTGACCCACTGAGAGCCGCCACCCCAACAGACAGAGGAGGAGGAGGAGG 486

Db 237 TCCGATTCACACGCTACACACCCACATTCGAACTTCGAAAGCCACACATCTCC 178

Qy 2466 TCGTTCATAGAGCCCGGTAAGTGAAGTTCACACCTAGAACACTGTCTCCGACCGGAGAAC 2525

Db 177 TGTCTTCATAGAGCCCGGTAAGTGAAGTTCACACCTAGAACACTGTCTCCGACCGGAGAAC 118

Qy 2526 GCGTCCCTTGGACCTGTATTTTACATGTGACTGTGCTTCTTCCCTGTGTCTTGAATG 2585

Db 117 GCGTCCCTTGGACCTGTATTTTACATGTGACTGTGCTTCTTCCCTGTGTCTTGAATG 58

Qy 2586 TTTAGACTCTTAAGAT 2601

Db 57 TTTAGACTCTTAAGAT 42

RESULT 8

LOCUS BX380892 1010 bp mRNA linear EST 08-MAY-2003

DEFINITION BX380892 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

ACCESSION BX380892

VERSION BX380892.1 GI:30439629

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 1010)

AUTHORS Pull-length cDNA libraries and normalization

TITLE Unpublished (2001)

JOURNAL

COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr; Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9799.r For more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODI058CE090P1&cluster=9799.r. Contact: Feng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID: CSODI058CE090P1.

FEATURES

source

1..1010

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODI058Y17"

/rname="PLACENTA COT 25-NORMALIZED"

/rsize_type="PLACENTA COT 25-NORMALIZED"

/clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime and enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 34.5%; Score 900.2; DB 13; Length 1010;

Best Local Similarity 98.6%; Pred. No. 1.2e-212;

Matches 934; Conservative 5; Mismatches 4; Indels 4; Gaps 3;

Qy 1599 GAGCCAGAGAACCCAGTCTTCCGAATGCGGTGCTCAGGAGGATTAATAATCTCTTC 1658

Db 66 GATTCACAGACAGCCAGCTTCCGAATGCGGTGCTCAGGAGGATTAATAATCTCTTC 125

Qy 1659 TGTGCAAGAGACAGCTCCAGAAAAAAGAGACACAGTGAAGAGAGAGAGTGTGT 1718

Db 126 TGTGCAAGAGACAGCTCCAGAAAAAAGAGACACAGTGAAGAGAGAGAGTGTGT 185

Qy 1719 CCTTCGCGAGTGAAGACTCGGAGAGAGAGAGAGCTTGTGAGAGA-GCAGCGCTCTGT 1777

Db 186 CCTTCGCGAGTGAAGACTCGGAGAGAGAGAGAGCTTGTGAGAGAGAGAGCGCTCTGT 245

Qy 1778 GGGGAGGATTCACATTAATGATGATTAAGCCAGAAAAAGCTGTCTCCCTGCGCTT 1837

Db 246 GGGGAGGATTCACATTAATGATGATTAAGCCAGAAAAAGCTGTCTCCCTGCGCTT 305

Qy 1838 CACGTGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1897

Db 306 CACGTGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 365

Qy 1898 TCTGATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1957

Db 366 TCTGATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 425

Qy 1958 ATCAAGACACCGGCTGACCATCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017

Db 426 ATCAAGACACCGGCTGACCATCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485

Qy 2018 CTGCGCTGTGTCATCCCTGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2077

Db 486 CTGCGCTGTGTCATCCCTGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545

Qy 2078 ACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2137

Db 546 ACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605

Qy 2138 GCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2197

Db 606 GCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 665

Qy 2198 CAGCGACATCCACCCCTGTGACATGATGATGATGATGATGATGATGATGATGATGAT 2257

Db 666 CAGCGACATCCACCCCTGTGACATGATGATGATGATGATGATGATGATGATGATGAT 725

Qy 2258 CAGTTTTCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2317

Db 726 CAGTTTTCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 785

Qy 2318 GGAATCTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2377

Db 786 GGAATCTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 845

Qy 2378 CAGATGAGGAGATTTTTTTTTTTTTCATGATGATGATGATGATGATGATGATGATGAT 2437

Db 846 CAGATGAGGAGATTTTTTTTTTTTTCATGATGATGATGATGATGATGATGATGATGAT 903

Qy 2438 ACATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2496

Db 904 ACATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 963

Qy 2497 CACCTAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2543

Db 964 CACCTAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1010

RESULT 9

LOCUS AL514138 1201 bp mRNA linear EST 08-MAY-2003

DEFINITION AL514138 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone

ACCESSION AL514138

VERSION AL514138.2 GI:30464023

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 1201)

AUTHORS Pull-length cDNA libraries and normalization

TITLE Unpublished (2001)

JOURNAL

COMMENT On Feb 13, 2001 this sequence version replaced gi:12777632.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CL0BB005ZB1RP1&cluster=9799.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CL0BB005ZB1RP1.

FEATURES

source

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL0BB005ZB1"
/issue_type="NEUROBLASTOMA"
/clone_1ib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN

Query Match 33.1%; Score 868; DB 9; Length 1201;
Best Local Similarity 95.1%; Pred. No. 1.3e-204;
Matches 919; Conservative 7; Mismatches 37; Indels 3; Gaps 3;

65 GAGACTTATCCACGCGGAAACGCTCTAATATCTGAGTGTGTAATAATCAAGCTTG 124
160 GAGACTTATCCACGCGGAAACGCTCTAATATCTGAGTGTGTAATAATCAAGCTTG 219
125 TTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 184
220 TTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 279
185 AGAGGTTTGTGTCGACGAGGTTTGTGTTATTTAGATTTGAGGAAAGTGTCAAG 244
280 AGAGGTTTGTGTCGACGAGGTTTGTGTTATTTAGATTTGAGGAAAGTGTCAAG 339
245 CTTCAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 304
340 CTTCAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 399
305 AGGCAATCAAAACCCACGACGAAAGTTCAGGAAACATGGGGTTTCCAGAGCACTA 354
400 AGGCAATCAAAACCCACGACGAAAGTTCAGGAAACATGGGGTTTCCAGAGCACTA 459
365 TCGCCAAAGGAGGAGGCGGAGGAGGCGGAGGCTGACCACTGAGAGCGCCAC 424
460 TCGCCAAAGGAGGAGGCGGAGGAGGCGGAGGCTGACCACTGAGAGCGCCAC 519
425 AGCAGAGCTGGGCGCTGTCTCCGCGGCACTGAGAGGAGCCCAAGCCGCACTGAG 484
520 AGCAGAGCTGGGCGCTGTCTCCGCGGCACTGAGAGGAGCCCAAGCCGCACTGAG 579
485 TGAAGAGTTCCTGAGCAATTCGCGGCGCGCGGAGGAGGAGGAGGAGGAGGAGG 544
580 TGAAGAGTTCCTGAGCAATTCGCGGCGCGCGGAGGAGGAGGAGGAGGAGGAGG 639
545 AGGATTTCTGTAGAGCCAGCTCTGCGCGGCAAGAGCCGAGAGAGCTCCGAGGAG 604
640 AGGATTTCTGTAGAGCCAGCTCTGCGCGGCAAGAGCCGAGAGAGCTCCGAGGAG 699
605 GGTGTGAAAGCGCTTCTGAGACCAAGAGCGGCCCAAGTGTGCTTCAAGCTGTGAAG 664
700 GGTGTGAAAGCGCTTCTGAGACCAAGAGCGGCCCAAGTGTGCTTCAAGCTGTGAAG 759
665 AACGACCAAGCTCTTCTGAAAGGTAAGAGGAGGAGTACCAAGTGTGCACTCCGAGTA 724
760 AACGACCAAGCTCTTCTGAAAGGTAAGAGGAGGAGTACCAAGTGTGCACTCCGAGTA 819
725 GTGACGAGATGGCTGACCTTGAAGAGCTTCAGATGCGCTTCGACAGGAAGCGGAAC 784

DB 820 GTGACGAGATGGCTGACCTTGAAGAGCTTCAGATGCGCTTCGACAGGAAGCGGAAC 879
QY 785 AGGAGCCACTGAGAGGAGCCCTGAAAGGATCCAGATGCGCTCCGAGAGAGCGCGG 844
DB 880 AGGAGCCACTGAGAGGAGCCCTGAAAGGATCCAGATGCGCTCCGAGAGAGAGCGCGG 938
QY 845 AGGAGGTTCCGCGGAGTGTGAGGCTCCGAGGCGAGTACACTGTGAGAGGCGTCTTC 904
DB 939 AGGAGGTTCCGCGGAGTGTGAGGCTCCGAGGCGAGTACACTGTGAGAGGCGTCTTC 997
QY 905 CCAAGTACGAGAGAGCCCGAGAGAGATGATGAGGAGTGTGTCCTCCGAGGAGAGTACA 964
DB 998 CCAAGTACGAGAGAGCCCGAGAGAGATGATGAGGAGTGTGTCCTCCGAGGAGAGTACA 1056
QY 965 GAGAGAGTAAAGTGTGAGAGGAGGAGGCGCTCAGAGCATCAAGATGAGAGAGCTGAGACT 1024
DB 1057 RAGAGGTAATTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
QY 1025 TGGGCG 1030
DB 1117 CGAGCC 1122

RESULT 10

AL553387 1201 bp mRNA linear EST 31-MAY-2003
LOCUS AL553387 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1075L03 5-PRIME, mRNA sequence.

ACCESSION AL553387
VERSION AL553387.2 GI:31275201
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

COMMENT On Feb 15, 2001 this sequence version replaced gi:12893174.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1075CP020P1&cluster=9799.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CS0D1075CF020P1.

FEATURES

source

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1075L03"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_1ib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 33.1%; Score 863.6; DB 9; Length 1201;
Best Local Similarity 93.5%; Pred. No. 1.6e-203;
Matches 914; Conservative 19; Mismatches 33; Indels 12; Gaps 3;

1 CTCGTCGCGCGCTCCGCACTCCGCGCGCTTCGAGGAAATGCTGCGAGACCTTACAGGC 60
|||||

```

Db      57 CCGGGATCCGTCGCCCACTCCGGCGGTTCGGGGAAATGGCTGAGAGCCTAGAGCC 116
Oy      61 CTGCG-----GAGCTTACTCAACGGGAAACAGCTCTAGATTAATCTGAGTTGTGAAAT 114
Db      117 CTGCGCGCTGCGGAAATTAATCAACGGGAAACAGCTCTAGATTAATCTGAGTTGTGAAAT 176
Oy      115 ACGAAGCCTGTAACTCGTGAACAGTGGCTGACAAACAGTGTGTGTGAGCTGGCTGTCT 174
Db      177 ACGAAGCCTGTAACTCGTGAACAGTGGCTGACAAACAGTGTGTGTGAGCTGGCTGTCT 236
Oy      175 GCTTGAACCCAGAGGTTTGTCTGCGCAGGTTTGTGTGTTAATTAGATTTCAGGAAA 234
Db      237 GCTTGAACCCAGAGGTTTGTCTGCGCAGGTTTGTGTGTTAATTAGATTTCAGGAAA 296
Oy      235 AGTGTCCAAAGCTTTCAAGTGTGTGAGAGGTTATGACAGCAAAAGGCGAATGCGCATGAG 294
Db      297 AGTGTCCAAAGCTTTCAAGTGTGTGAGAGGTTATGACAGCAAAAGGCGAATGCGCATGAG 356
Oy      295 GAGGACCTTAAGGCAATCAAAACCCAGCAAAAGTTTCAAGGAAAATGAGGAGTTTTCGA 354
Db      357 GAGGACCTTAAGGCAATCAAAACCCAGCAAAAGTTTCAAGGAAAATGAGGAGTTTTCGA 416
Oy      355 AGGACCACTATGCGCAAGGAGAGGCGGAGGAGAGCGCGAGGCTGACCCACTGAGCCG 414
Db      417 AGGACCACTATGCGCAAGGAGAGGCGGAGGAGAGCGCGAGGCTGACCCACTGAGCCG 476
Oy      415 CCAACCCCAACAGCAGAGCTGGGCTGTCTCTGCGCGGAGTGGAGAGGCAAGCCCAAGCC 474
Db      477 CCAACCCCAACAGCAGAGCTGGGCTGTCTCTGCGCGGAGTGGAGAGGCAAGCCCAAGCC 536
Oy      475 ACTGAGCGCGTGGAGAGTTCCTGACCAATTTGCGCGCGCGCGGAGAGAGAGCATGCT 534
Db      537 ACTGAGCGCGTGGAGAGTTCCTGACCAATTTGCGCGCGCGCGGAGAGAGAGCATGCT 596
Oy      535 GTCTCCCTGAGAGATTCTGTGAGCCCAAGTCCTGCGCGGAGAGAGAGAGAGAGAGCC 594
Db      597 GTCTCCCTGAGAGATTCTGTGAGCCCAAGTCCTGCGCGGAGAGAGAGAGAGAGAGCC 656
Oy      595 TCCGAGGCGAGGTGAGAAAGCCCTTCTGAGACCAAGAGGCGCCCAAGTCCTTCCACA 654
Db      657 TCCGAGGCGAGGTGAGAAAGCCCTTCTGAGACCAAGAGGCGCCCAAGTCCTTCCACA 716
Oy      655 GCTGTGAAGAGAGAGAGAGAGCTCTTCTGAAAGAGTGAAGAGAGAGAGAGAGAGAG 714
Db      717 GCTGTGAAGAGAGAGAGAGAGCTCTTCTGAAAGAGTGAAGAGAGAGAGAGAGAGAG 776
Oy      715 AACTCCGATAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774
Db      777 AACTCCGATAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836
Oy      775 AAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 834
Db      837 AAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896
Oy      835 AAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 894
Db      897 AAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 956
Oy      895 GAGCGTCTGCGCAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954
Db      956 GAGCGTCTGCG-----ARTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1010
Oy      955 AAAGATGACAGAGAGAGT 972
Db      1011 AAAGATGACAGAGAGAGT 1028

```

```

RESULT 11
AL550675/c 1018 bp mRNA linear EST 31-MAY-2003
LOCUS AL550675 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS001056YK04 3-PRIME, mRNA sequence.
ACCESSION AL550675

```

```

VERSION AL550675.2 GI:31272492
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE li W.B., Gruber C., Jesse J., and Polayes D.
AUTHORS Full-length cDNA libraries and normalization
TITLE Unpublished (2001)
JOURNAL On Feb 15, 2001 this sequence version replaced gi:12887874.
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS001056BF02NP1&cluster=9799.r. Contact :
Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CS001056BF02NP1.
FEATURES
source
1..1018
Location/Qualifiers
1..1018
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS001056YK04"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

```

Query Match 32.4%; Score 845.2; DB 9; Length 1018;
Best Local Similarity 97.5%; Pred. No. 5,9e-199;
Matches 856; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

```

```

Oy      1189 AGAATGAGGAAAGACTATATATCTGCCCAATGCAATTCGACATTTCTGCAAGTGCAGATGACT 1248
Db      879 AAGGAATGGAGAGACTATATCTGCGCAATGCAATTCGACATTTCTGCAAGTGCAGATGACT 820
Oy      1249 CATTGAGAAACGGCAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308
Db      819 CATTGAGAAACGGCAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 760
Oy      1309 GATTGACAAAGTATAGGAAACAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368
Db      759 GATTGACAAAGTATAGGAAACAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 700
Oy      1369 AGAATGAGAAAGCTGCAATTCGCAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428
Db      699 AGAATGAGAAAGCTGCAATTCGCAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 640
Oy      1429 ATAGAGGCGCTGCTGCTCAAAATGATATGGCCCCCGGTGTGTCACGTCGCGAGCCC 1488
Db      639 ATAGAGGCGCTGCTGCTCAAAATGATATGGCCCCCGGTGTGTCACGTCGCGAGCCC 580
Oy      1489 GACTCGGTGTACTGAGAGTATGATCTGTATCTCAACACGCGCGAGAGAGAGAGAGAG 1548
Db      579 GACTCGGTGTACTGAGAGTATGATCTGTATCTCAACACGCGCGAGAGAGAGAGAGAG 520
Oy      1549 CTAAGCTCAGGTAAAGACAGAGGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
Db      519 CTAAGCTCAGGTAAAGACAGAGGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 460
Oy      1609 AAGCCAGCTTCGGAATAGCGGTGCTCAGCGAGGATTTAAATCTCTTGTGCAAG 1668
Db      459 AAGCCAGCTTCGGAATAGCGGTGCTCAGCGAGGATTTAAATCTCTTGTGCAAG 400

```

QY 1669 AGACCACTCCAGAAAAAGAGACCACTGAAAGAGAGAGTGTGTCTCTGCGCG 1728
| | | | |
Db 399 AGACCACTCCAGAAAAAGAGACCACTGAAAGAGAGAGTGTGTCTCTGCGCG 340
| | | | |
QY 1729 AGTGAACACTCGGGAAGAGAGAGCTTGTGAGAGACACCGCGTCTGTGGCGAGCAT 1788
| | | | |
Db 339 AGTGAACACTCGGGAAGAGAGAGCTTGTGAGAGACACCGCGTCTGTGGCGAGCAT 280
| | | | |
QY 1789 CACAAATTACAAATGAGTAAAGCAAGAAAGATGCTGTCTCTGCGCGTCACTGTAT 1848
| | | | |
Db 279 CACAAATTACAAATGAGTAAAGCAAGAAAGATGCTGTCTCTGCGCGTCACTGTAT 220
| | | | |
QY 1849 AATGTATGATCACTGAGGGTGTGCTCTGAGACCCCTCCGTTCTTTCTGATAGCC 1908
| | | | |
Db 219 AATGTATGATCACTGAGGGTGTGCTCTGAGACCCCTCCGTTCTTTCTGATAGCC 160
| | | | |
QY 1909 ATCCCTGAGGCTGTCCAGAGCTGGAGTTGACGTTGTATTAAGTATCAAGAC 1968
| | | | |
Db 159 ATCCCTGAGGCTGTCCAGAGCTGGAGTTGACGTTGTATTAAGTATCAAGAC 100
| | | | |
QY 1969 CGGCTGACCAATGACGCGGGAAGAGAGCCCATGTCAGATGCCCTGCGCCGTGT 2028
| | | | |
Db 99 CGGCTGACCAATGACGCGGGAAGAGAGCCCATGTCAGATGCCCTGCGCCGTGT 40
| | | | |
QY 2029 CCATCCCTAGTGTGACGAGCTCTCTCTCACTGTTTC 2066
| | | | |
Db 39 CCATCCCTAGTGTGACGAGCTCTCTCTCACTGTTTC 3
| | | | |
RESULT 12
LOCUS BX404270 1050 bp mRNA linear EST 15-MAY-2003
DEFINITION BX404270 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CU08B022ZG10 5-PRIME, mRNA sequence.
ACCESSION BX404270
VERSION BX404270.1 GI:30764484
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1050)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CU08B022ZG10RPLcluster=9799.r. Contact :
Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CU08B022ZG10RPL.
Location/Qualifiers
FEATURES
SOURCE
1. 1050
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CU08B022ZG10"
/cissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT 6, 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
ORIGIN
Query Match 32.1%; Score 838.2; DB 13; Length 1050;

Best Local Similarity 96.7%; Pred. No. 3.3e-197;
Matches 875; Conservative 12; Mismatches 10; Indels 8; Gaps 3;
QY 2 TCGGTGGCGTCCCGCCACTCCGCGCGCTTCGCGGAAATGCTGCGAGACCTTAGAGCC 61
| | | | |
Db 52 TCCCGGAGTGTCCCGCCACTCCGCGCGCTTCGCGGAAATGCTGCGAGACCTTAGAGCC 111
| | | | |
QY 62 TCGC-----GAGCTTACTCCAGCGGAAAGAGCTCTAGATAATGAGTTGTAAATA 115
| | | | |
Db 112 TCGCGCTCGCGGASTTACTCCAGCGGAAAGAGCTCTAGATAATGAGTTGTAAATA 171
| | | | |
QY 116 CGAAGCTGTATCTGTGAAAGATGAGCTGACCAAGCTGTGTGTGAGCTGTCTG 175
| | | | |
Db 172 CGAAGCTGTATCTGTGAAAGATGAGCTGACCAAGCTGTGTGTGAGCTGTCTG 231
| | | | |
QY 176 CTGGAACCAAGAGTTGTGTCTGACGAGGTTTGTGTATTTAGATTTAGAGGAAAA 235
| | | | |
Db 232 CTGGAACCAAGAGTTGTGTCTGACGAGGTTTGTGTATTTAGATTTAGAGGAAAA 291
| | | | |
QY 236 GTGTCCAAAGCTTTCAGTGTGAGAGGATGACGACCAAGGCGACCCGACATGAG 295
| | | | |
Db 292 GTGTCCAAAGCTTTCAGTGTGAGAGGATGACGACCAAGGCGACCCGACATGAG 351
| | | | |
QY 296 AGGCACTTAAGGCTTCAAAACCAACCAAGATTTAGAGAAACATGGGGTTTGGAA 355
| | | | |
Db 352 AGGCACTTAAGGCTTCAAAACCAACCAAGATTTAGAGAAACATGGGGTTTGGAA 411
| | | | |
QY 356 GGAACCACTATCCCAAGGAGAGAGGCGAGGGAAGCGGAGGCTGACCCACTGAGCCGC 415
| | | | |
Db 412 GGAACCACTATCCCAAGGAGAGAGGCGAGGGAAGCGGAGGCTGACCCACTGAGCCGC 471
| | | | |
QY 416 CACCCCAACAGAGAGCTGGGCTGTCTCTGCGCGCGAGTGGAGAGCAAGCCCAAGCGCA 475
| | | | |
Db 472 CACCCCAACAGAGAGCTGGGCTGTCTCTGCGCGCGAGTGGAGAGCAAGCCCAAGCGCA 531
| | | | |
QY 476 CTGAGCGGTGAGAGAGTTCTGACCAATTGCGCGCGCGCGAGAGAGAGATGCTG 535
| | | | |
Db 532 CTGAGCGGTGAGAGAGTTCTGACCAATTGCGCGCGCGCGAGAGAGAGATGCTG 591
| | | | |
QY 536 TCTCCCTGAGAGATCTGTGTGAGAGCCAGGTCCTGCCCCGCGACAGAGCGAGACAGCT 595
| | | | |
Db 592 TCTCCCTGAGAGATCTGTGTGAGAGCCAGGTCCTGCCCCGCGACAGAGCGAGACAGCT 651
| | | | |
QY 596 CCGAGGCGAGCTGTGAAAGCGCTTCTGAGACCAAGAGCGGCGCCAGTGTCTTCCAGG 655
| | | | |
Db 652 CCGAGGCGAGCTGTGAAAGCGCTTCTGAGACCAAGAGCGGCGCCAGTGTCTTCCAGG 711
| | | | |
QY 656 CTGTGAAGAGAGAGACCAAGCTCTTCTGAAAGGTGAAAGAGAGGAGTACCAAGATGAC 715
| | | | |
Db 712 CTGTGAAGAGAGAGACCAAGCTCTTCTGAAAGGTGAAAGAGAGGAGTACCAAGATGAC 771
| | | | |
QY 716 CCTCGATAGTGAACAGCGATGAGCTGACCTTGAAAGAGCTTGAGATGCGCTTCCAGGA 775
| | | | |
Db 772 CCTCGATAGTGAACAGCGATGAGCTGACCTTGAAAGAGCTTGAGATGCGCTTCCAGGA 831
| | | | |
QY 776 ACGCGGAAACAGAGGCCACTGAGAGGCCCTTGAAGAGGATTCAGAGTGCCTTCGGAAG 835
| | | | |
Db 832 ACGCGGAAACAGAGGCCACTGAGAGGCCCTTGAAGAGGATTCAGAGTGCCTTCGGAAG 890
| | | | |
QY 836 ACGCGCGGAGAGAGGCTCCCGCGAGACTGTGGGCTTCCAGAGCTCACTGATGACCTGTGAGG 895
| | | | |
Db 891 ACGCGCGGAGAGAGGCTCCCGCGAGACTGTGGGCTTCCAGAGCTCACTGATGACCTGTGAGG 949
| | | | |
QY 896 GCGTC 900
| | | | |
Db 950 GCGTC 954
| | | | |
RESULT 13
LOCUS CD364822/c 850 bp mRNA linear EST 29-MAY-2003
DEFINITION UI-H-F12-bjm-d-20-0-UI.s1 NCI CGAP F12 Homo sapiens cDNA clone
UI-H-F12-bjm-d-20-0-UI 3', mRNA sequence.

ACCESSION CD364822
 VERSION GI:31148912
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 850)
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: Dr. Gary W. Hunninghake, U of I
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/cgap.html>
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source
 location/Qualifiers
 1..850
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FT2-bjm-d-20-0-UI"
 /issue_type="Alveolar Macrophage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_id="NCI CGAP FT2"
 /note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a
 modified polylinker; Site_1: Ecor I; Site_2: Not I;
 NCI CGAP FT2 is a subcloned cDNA library constructed from
 a pool of 81 RNA samples from Alveolar Macrophages
 challenged with different treatments. The library was
 subcloned according to Bonaldi, Lemon and Soares, Genome
 Research, 6:791-806, 1996. The tissue was provided by Dr.
 Gary W. Hunninghake of the University of Iowa.
 TAG LIB=UI-H-FT2
 TAG TISSUE=Human Lung Alveolar Macrophage
 TAG_SEQ=GCCCATGCCG"

ORIGIN
 Query Match 31.8%; Score 830.4; DB 14; Length 850;
 Best Local Similarity 99.6%; Pred. No. 2.7e-195;
 Matches 831; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1258 ACCGCAATCGACGAGAGCTTAATGAGACCTGAGAGTGTGAGCAACGATTTGACA 1317
 Db 850 ACGGCAATCGACGAGAGCTTAATGAGAGCTGAGAGTGTGAGCAACGATTTGACA 791

1318 AGTATGAGCAATATGAGAGCTGAGAGCAAGCAAGGATTAAGGATTTGAG 1377
 Db 790 AGTATGAGCAATATGAGAGCTGAGAGCAAGCAAGGATTAAGGATTTGAG 731

1378 AAAGCTGCAATTCAGAGTGGCAAGAACTCAAGATTTTCAGCTGTGTAGAGCG 1437
 Db 730 AAAGCTGCAATTCAGAGTGGCAAGAACTCAAGATTTTCAGCTGTGTAGAGCG 671

1438 CCTGTGCTCAAAATGATTTGGCCCGGCTTTCAGCTGTGCGCAACCGGACTCGGTG 1497
 Db 670 CCTGTGCTCAAAATGATTTGGCCCGGCTTTCAGCTGTGCGCAACCGGACTCGGTG 611

1498 TACTGCAATTAATGATCTGATCTCAACACGCGCGAGCAATGAAGTTTCTAAGCTCA 1557
 Db 610 TACTGCAATTAATGATCTGATCTCAACACGCGCGAGCAATGAAGTTTCTAAGCTCA 551

1558 GGTAAAGAAAGCAAGCCAAAGCTTAAGAAAGATGAAGATGAGCCAGAGAGCCCACT 1617
 Db 550 GGTAAAGAAAGCAAGCCAAAGCTTAAGAAAGATGAAGATGAGAGAGCCCACT 491

1618 CTTCCGAATTCGAGTCTCAGGACGATTAATAATCTTCTGTGCAAGAGACCACT 1677
 Db 490 CTTCCGAATTCGAGTCTCAGGACGATTAATAATCTTCTGTGCAAGAGACCACT 431

1678 CCGAGAAAAAAGAGACCAAGTGAAGAAAGGACAGTGTGTCTCCCGGAGTGAAGCA 1737
 Db 430 CCGAGAAAAAAGAGACCAAGTGAAGAAAGGACAGTGTGTCTCCCGGAGTGAAGCA 371

1738 CTGGGAAAGAGAGAGCTTGTGAAGAGACAGCCGCTGTGGGAGAGCATCAATTAAC 1797
 Db 370 CTGGGAAAGAGAGAGCTTGTGAAGAGACAGCCGCTGTGGGAGAGCATCAATTAAC 311

1798 AATGCAATTAAGCCAGAAAGACAGTCTCTCCCTGCGCTCACTGTTATTAATGATG 1857
 Db 310 AATGCAATTAAGCCAGAAAGACAGTCTCTCCCTGCGCTCACTGTTATTAATGATG 251

1858 TATCACTTGAAGGAGTGTGCTCTCTGAGACCCCTTCCTTTCTGTGATAGCATTCCTCG 1917
 Db 250 TATCACTTGAAGGAGTGTGCTCTCTGAGACCCCTTCCTTTCTGTGATAGCATTCCTCG 191

1918 GCGTGTCCAGAGCTGGGAGTGTGAGCTTGTGTTAAGCTGATCAGACACCGGCTGAC 1977
 Db 190 GCGTGTCCAGAGCTGGGAGTGTGAGCTTGTGTTAAGCTGATCAGACACCGGCTGAC 131

1978 CATCAGCGGAGAGAGAGCCCATATGTCAGATGCTCTCTGCTGCGCTGTTCATTCCTTA 2037
 Db 130 CATCAGCGGAGAGAGAGCCCATATGTCAGATGCTCTCTGCTGCGCTGTTCATTCCTTA 71

2038 GTCTGTCCAGAGCTTCTCTGCTCACTGTTTCCAAAGCTGTAACTTCACTGTGTAA 2091
 Db 70 GTCTGTCCAGAGCTTCTCTGCTCACTGTTTCCAAAGCTGTAACTTCACTGTGTAA 17

RESULT 14
 AK014422 1614 bp mRNA linear HTC 20-SEP-2003
 LOCUS AK014422
 DEFINITION Mus musculus 18 days pregnant adult female placenta and extra
 embryonic tissue cDNA, RIKEN full-length enriched library,
 clone:3830408B01 product:death inducer-oblierator-1, full insert
 sequence.

ACCESSION AK014422
 VERSION AK014422.1 GI:12852263
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20493374
 PUBMED 11042159

REFERENCE 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Ozawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Tanaka, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichannel sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913

PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS (bases 1 to 1614)
Adachi, J., Alizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arahata, T., Bono, H., Carninci, P., Fukuda, S., Fukushima, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Iwaw, M., Kanakawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Konda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Shukri, H., Tagami, M., Tagawa, A., Takahashi, P., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Yamatsuta, M., and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda PLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.
FEATURES
SOURCE 1..1614
Location/Qualifiers
1..1614
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:3830408B01"
/db_xref="MGI:1910351"
/db_xref="taxon:10090"
/clone="3830408B01"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18 days pregnant adult"
331..1614
/note="deach inducer-oblierator-1 (MGD|MGI:1344352, GBI|W238332, evidence: BLASTN, 99%, match=2599) putative"
ORIGIN
Query Match 31.3%; Score 817.8; DB 11; Length 1614;
Basic Local Similarity 73.1%; Pred. No. 4.4e-192;
Matches 1127; Conservative 0; Mismatches 387; Indels 27; Gaps 5;

QY 21 TCCGCGGCGCTTCGAGAAATGGCTGCGAGACCTTAGAGGCTGCGAGCTACTCCACGG 80
DB 100 TCCGCTGTAAGCTCTGAAATAGCTGCGGATCCCGCGGCCGGGAGCTTGTTAGAGG 159
QY 81 GAACAGCCTCTAGATATCTGAGTTGTGAAATAGAAAGCTGTACTGTGAAACAGTG 140
DB 160 CAGTCCCACTGTCCCTTGTGTGTGAAAGCTCCGGAATCTTCTCATGTATGAACTGTG 219
QY 141 GCTGACAACTGTGTGTGTGAGCCTGCTGTCTGTGTGAGCCAGGAGTTCTGTCC 200
DB 220 ACTGACAACTAGTGGGTGAG--GCTTGGCCGTCTGCTTAC--TGAGCC 265
QY 201 AGGCTTTTGTGTGTGTAGATTT--CAGGAAAGTGTCCAGCTTCAGTGTGAG 259
DB 266 CAGGCTTAATTTATGTAGGATTTCCAGCAAGTTTCAAGTTTCAAGTTTGGGA 325
QY 260 CAGTATGAGACAAAGCGCAGCCGAGCAATGAGAGGCACTTAAGGCATCAACCCA 319
DB 326 CAGGTATGATATTAAGGCACTTGAACATGAGAGACATCCAGCTATCAAAACCA 385
QY 320 CCAGCAAGATTCAGAAACATGAGGTTTTCAGAGCACTATGCGCAAGCAGAG 379
DB 386 CAGTATGAGATTCAGAAACCTGAGGTTTTCAGAAACCAAGATTCAGAG 445
QY 380 GCGCAGGAGGACCGGAGGCTGACCTGAGACCCGACAGCAAGCAGAGCTGGCC 439
DB 446 GTGCGAGAGACAGAGGTGAGCCCACTGAGCAGCAAC-----ACAGCAGCATPACC 499
QY 440 TGTCCCTGCGCGCAGTGGAGGAGCCCAAGCAGCACTGAGCGGTGAGAGCTTCTGA 499
DB 500 TTTCTCTGCGCGCAGTGGAGGAGCCCAAGCAGCACTGAGCGGTGAGAGCTTCTGA 559
QY 500 CCATTGCGCGCGCGCGCAGAGAGAGCAGTCTGTCTCTCTGAGAGATTCGTGAGC 559
DB 560 CCAAGCTTCGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619
QY 620 CCACATCTTCCACAGTCACTGATGTGAGACAGCTTCCAGAGGAGAGAGAGAGAG 679
DB 620 CTGAGACGAGAGAGCGCCCGCAGTCTGCTTCCAGAGCTGTGAGAGAGCAGCTCTT 679
QY 680 CTGAAATCAAAATGCTCTGTA-----TGTACTCTTGAAGAGAACTCTGCTCTT 733
DB 734 CTGAAATGCAAAAG 793
QY 740 TGAATCTGAAAGCTTCAAAATGCTCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 799
DB 794 TTAAGTTGAGGAGATTCAGAAACCGCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 853
QY 800 GCGCCCTGAAAGAGATTCAGAGTGTGCTGCGAGAGAGAGAGAGAGAGAGAGAG 859
DB 854 GCTTCTTGAAG 913
QY 860 AGACTGTGAGCTCCAGAGCAGTGTGACACTGTGAGAGAGAGAGAGAGAGAGAGAG 919
DB 914 AATCTGAGAGTGTCCAAATAGCAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAG 973
QY 920 CCGAGAGAGATTCAG 979
DB 974 CTGAGAGTATGTCAG 1033
QY 980 AGGGAAG 1039
DB 1034 AAGGGAAG 1093
QY 1040 CTGATGTGAG 1099
DB 1094 CTGATGTGAG 1153
QY 1100 ACAAGTTATGATTTGCTGTGACCCCTGTGAAAGATGTTTCAAGAGAGAGAGAGAG 1159

Db	1154	ACAGGTTTATGATCTGCTGTGATGCGTGTAGAGAGTGGTTCCATAGTGATCTGTGGGTA	1213
Qy	1160	TTTTCTGAGGCTCGAGGAGGCTTTTGGAAAGAAATGGGAAGACTAATATCTGCCAACT	1219
Db	1214	TTTTCTGAGGCCCCGAGGGGGGCTCTCTGGAAAGAAACGGGGAAGACTACATCTGCCCAATT	1273
Qy	1220	GCACCAATTTCTGCAAGTGCAGATATGACTATTCAGAAACGGCGAGATCAGAGGAAGCTA	1279
Db	1274	GCACCAATTTTTCAGAGTGCAGATGATGAAACGGTAGCGCCACCGATATGAGGACTCTTG	1333
Qy	1280	AATGAGAGCTCGAGATGCTGATGGACCAATTTGTAACAATATAGAAACAATATAGCAGA	1339
Db	1334	GGTGCAAGTCTGTGGGTGCTGATATGACACAGACTGCACAAAGCATAGGACATAGACAGA	1393
Qy	1340	AGCTTAGCGAAGACCAAGGAGTAAAGGTTAGAAATTGAGAAAGCTGCAAATCCAAGTGC	1399
Db	1394	AGTCGAGGAAGACCAAGGCAATTAAGGTTAGAAATTGAGAAAGCCACAGCCGCA	1453
Qy	1400	AGAGAAATCTCAAGATCTTCCAGCTGTGTGATAGAGGCGCTGTGTGCTCAAAATGATATG	1455
Db	1454	AAAAAAATCTCAAGATATTTCCAGCTGTGTGTAGAGGCTCTGTGTCTCTAAATGCAATTG	1513
Qy	1460	GCCCCGGGTGCTGACGTTGGGCGACGCCGACTGGGTGTACGCGAGTATGACTGTATTC	1519
Db	1514	GCCCTGGGTGTTCCAGGTGTACACAGCTGTGACTCTGTATATTCGAGATATGATGCTGATTC	1573
Qy	1520	TCAACACGCGCAGCGACATGAAATTTCTAAGCTCAGGT	1560
Db	1574	TCAACACGCGACGACTTCATGAGATTTCTAAGTTCAGGT	1614

RESULT	15
LOCUS	AL550431
DEFINITION	1119 bp mRNA linear EST 31-MAY-2003
ACCESSION	AL550431 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CSODI057YC01 5-PRIME, mRNA sequence.
VERSION	AL550431
KEYWORDS	AL550431..2 GI:31272248
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Li W.B., Gruber C., Jessee J. and Polayes D. Full-length cDNA libraries and normalization Unpublished (2001) On Feb 15, 2001 this sequence version replaced gi:12867402.
AUTHORS	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France
TITLE	Email: segre@genoscope.cns.fr Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9799.r For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CSODI057AB01P&cluster=9799.r . Contact :
JOURNAL	Peng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSODI057AB01P.
COMMENT	Location/Qualifiers 1..1119 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSODI057YC01" /issue_type="PLACENTA COT 25-NORMALIZED" /clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED" /note="First strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 23, 2004, 07:35:03 ; Search time 192.11 Seconds
(without alignments)
8281.924 Million cell updates/sec

Title: US-09-787-016A-2
Perfect score: 2867

Sequence: 1 tccgtgtagctctgtaaat.....tttcttctctgtctgctg99 2867

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/Backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	77.8	2.7	7218	1	US-08-232-463-14
C 2	55.4	1.9	6256	4	US-09-620-312D-269
C 3	40.2	1.4	7218	1	US-08-232-463-14
C 4	37.8	1.3	505	4	US-09-621-976-15639
C 5	37.8	1.3	176373	3	US-09-128-155-17
C 6	37.2	1.3	6755	3	US-08-931-999-4
C 7	36.6	1.3	248	3	US-09-007-005-32
C 8	36.6	1.3	248	3	US-09-244-796-32
C 9	36.6	1.3	277	3	US-09-007-005-3
C 10	36.6	1.3	277	3	US-09-244-796-3
C 11	36.4	1.3	765	4	US-09-328-352-3047
C 12	36	1.3	474	4	US-09-621-976-18033
C 13	35.8	1.2	1492	4	US-09-196-520-7
C 14	35.6	1.2	289	3	US-09-007-005-17
C 15	35.6	1.2	289	3	US-09-244-796-17
C 16	35.6	1.2	2244	3	US-09-061-764A-18
C 17	35.6	1.2	2824	4	US-09-016-434-1349
C 18	35	1.2	190	3	US-09-188-930-57
C 19	35	1.2	190	4	US-09-312-283C-57
C 20	35	1.2	190	4	US-09-312-283C-402
C 21	35	1.2	472	4	US-09-621-976-18485
C 22	34.8	1.2	1230025	4	US-09-198-452A-1
C 23	34.6	1.2	245	1	US-08-139-937-8
C 24	34.6	1.2	245	5	PCT-US93-11310-8
C 25	34.4	1.2	399	4	US-09-621-976-8976
C 26	34.2	1.2	1501	3	US-09-196-520-1
C 27	34	1.2	732	4	US-09-134-000C-348

C 28	34	1.2	1146	1	US-08-487-810-1	Sequence 1, Appl1
C 29	33.8	1.2	505	4	US-09-621-976-15639	Sequence 15639, A
C 30	33.6	1.2	424	4	US-09-621-976-7975	Sequence 7975, Ap
C 31	33.6	1.2	1886	4	US-09-594-506-31	Sequence 31, Appl
C 32	33.6	1.2	2278	4	US-09-620-312D-1003	Sequence 1003, Ap
C 33	33.6	1.2	2278	4	US-09-148-545-78	Sequence 78, Appl
C 34	33.6	1.2	2278	4	US-09-148-545-131	Sequence 131, App
C 35	33.4	1.2	1864	4	US-09-205-258-225	Sequence 225, App
C 36	33.4	1.2	5066	4	US-08-961-527-79	Sequence 79, Appl
C 37	33.4	1.2	5345	1	US-08-044-618-7	Sequence 7, Appl1
C 38	33.4	1.2	11049	4	US-10-204-708-23	Sequence 23, Appl1
C 39	33.4	1.2	202001	4	US-09-734-674-3	Sequence 45, Appl1
C 40	33.4	1.2	1497	4	US-09-489-039A-6080	Sequence 6080, Ap
C 41	33.2	1.2	1950	4	US-09-614-891-2	Sequence 2, Appl1
C 42	33.2	1.2	2135	4	US-08-933-711B-17	Sequence 17, Appl1
C 43	33.2	1.2	2501	4	US-09-614-891-3	Sequence 3, Appl1
C 44	33.2	1.2	364	4	US-09-621-976-17202	Sequence 17202, A
C 45	33	1.2				

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, P.
APPLICANT: SCHRIPLINGER, P.
APPLICANT: FALKNER, P. G. F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: BP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)836-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9DC-F18
US-08-232-463-14
Query Match 2.7%; Score 77.8; DB 1; Length 7218;

APPLICATION NUMBER: US 08/7710,561
FILING DATE: 19-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 25043-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 816/474-9950
TELEFAX: 816/474-9057
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6755 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: *Staphylococcus aureus*
STRAIN: UT0007

Query Match	1.2%	Score 35.8	DB 3	Length 1492
Best Local Similarity	46.9%	Pred. No. 2,2		
Matches 112	Conservative 0	Mismatches 127	Indels 0	Gaps 0

QY	1381	CCTGTCGTAGAGGCTCTGTCGTCCTTAATGACATTGGCCCTGGGTGTCAGTGTACGA	1440
Db	786	CTTGAGGAAAAGTCGTCGTCGTTGTTCCATGATATATTAGCTTATGTATGCGAGAACCTT	727

QY	1441	CAGCTGACTCTGTGTATTTGCAGTAAATGACTGTTCCTCAACAGGACGACGTACACTG	1500
Db	726	CTCCCTGCATTGGCTGGTTGGCATTAATGAAAGCAACTCTGCACTTCCAACTGTGCTTG	667

QY	1501	AGATTTCTAAAGTTGAGTAAAGAACAAAAACAAACCCAAAGAAAGGTCAAGACGAAG	1560
Db	666	GCGGGTATATATTTTCAAACTTCCACAAACCCCTCAACACTGATGTCAAGGCCACT	607

QY	1561	CCAGAAAAGTCAGTCTTCCAAATGACAGTGTTCAGGTGGGATTAANAATCTCTTCTGT	1619
Db	606	ATACTCTCTCAGATATTTTCAACGAGTTGTCAAGGTTCAATCATATATCTAGTCT	548

RESULT 14
US-09-007-005-17
? Sequence 17, Application US/09007005B

RESULT 15-796-17
 US-09-244-796-17
 : Sequence 17, Application US/09244796
 : Patent No. 6281344
 : GENERAL INFORMATION:
 : APPLICANT: Szostak, Jack W.
 : APPLICANT: Roberts, Richard W.
 : APPLICANT: Liu, Rhee
 : TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
 : TITLE OF INVENTION: FUSIONS
 : FILE REFERENCE: 00786/350007
 : CURRENT APPLICATION NUMBER: US/09/244,796
 : CURRENT FILING DATE: 1999-02-05
 : EARLIER APPLICATION NUMBER: 60/035,963
 : EARLIER FILING DATE: 1997-01-27
 : EARLIER APPLICATION NUMBER: 60/064,491
 : EARLIER FILING DATE: 1997-11-06
 : EARLIER APPLICATION NUMBER: 09/007,005
 : EARLIER FILING DATE: 1998-01-14
 : NUMBER OF SEQ ID NOS: 33

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 14:17:01 ; Search time 1363.62 Seconds

(Without alignments)
9479.366 Million cell updates/sec

Title: US-09-787-016a-2

Perfect score: 2867

Sequence: 1 tccgtgtagctcgtgaat.....cttgcctctgctgctg 2867

Scoring table: IDENTITY NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	558	19.5	1764	9	US-09-925-301-327 Sequence 327, App
2	354.6	12.4	645	15	US-10-029-386-25091 Sequence 25091, A
3	288	10.0	478	10	US-09-918-995-19127 Sequence 19127, A
4	117.8	4.1	536	15	US-10-029-386-11381 Sequence 11381, A
5	65	2.3	65	10	US-09-908-975-25239 Sequence 25239, A
6	57.8	2.0	915	9	US-09-764-864-471 Sequence 471, App
7	57.8	2.0	2320	15	US-10-175-523-142 Sequence 142, App
8	55.4	1.9	5975	10	US-09-873-367C-112 Sequence 112, App
9	55.4	1.9	6256	15	US-10-037-270-269 Sequence 269, App
10	55.4	1.9	6256	16	US-10-117-722-269 Sequence 269, App
11	55.4	1.9	6948	10	US-09-930-213-315 Sequence 315, App
12	50.4	1.8	653	15	US-10-184-644-402 Sequence 402, App
13	50.4	1.8	653	15	US-10-184-634-402 Sequence 402, App
14	48	1.7	469	13	US-10-085-783A-56748 Sequence 56748, A

15	48	1.7	469	16	US-10-242-535A-56748	Sequence 56748, A
16	48	1.7	483	9	US-09-864-761-2140	Sequence 2140, App
17	47.4	1.7	594	13	US-10-142-426-10	Sequence 10, App1
18	47.4	1.7	594	15	US-10-123-155-10	Sequence 10, App1
19	47.4	1.7	594	15	US-10-146-731-10	Sequence 10, App1
20	47.4	1.7	594	15	US-10-140-472-10	Sequence 10, App1
21	47.4	1.7	594	15	US-10-141-761-10	Sequence 10, App1
22	47.4	1.7	594	15	US-10-142-885-10	Sequence 10, App1
23	47.4	1.7	594	15	US-10-158-790-10	Sequence 10, App1
24	47.4	1.7	594	16	US-10-137-871-10	Sequence 10, App1
25	47.4	1.7	594	16	US-10-140-923-10	Sequence 10, App1
26	47.4	1.7	594	16	US-10-141-756-10	Sequence 10, App1
27	47.4	1.7	594	16	US-10-141-759-10	Sequence 10, App1
28	47.4	1.7	594	16	US-10-140-805-10	Sequence 10, App1
29	47.4	1.7	594	16	US-10-140-864-10	Sequence 10, App1
30	42.8	1.5	114615	13	US-10-087-192-676	Sequence 676, App
31	41.4	1.4	745	13	US-10-142-426-68	Sequence 68, App1
32	41.4	1.4	745	15	US-10-123-155-68	Sequence 68, App1
33	41.4	1.4	745	15	US-10-146-731-68	Sequence 68, App1
34	41.4	1.4	745	15	US-10-140-472-68	Sequence 68, App1
35	41.4	1.4	745	15	US-10-141-761-68	Sequence 68, App1
36	41.4	1.4	745	15	US-10-142-885-68	Sequence 68, App1
37	41.4	1.4	745	15	US-10-158-790-68	Sequence 68, App1
38	41.4	1.4	745	16	US-10-137-871-68	Sequence 68, App1
39	41.4	1.4	745	16	US-10-140-923-68	Sequence 68, App1
40	41.4	1.4	745	16	US-10-141-756-68	Sequence 68, App1
41	41.4	1.4	745	16	US-10-141-759-68	Sequence 68, App1
42	41.4	1.4	745	16	US-10-140-805-68	Sequence 68, App1
43	41.4	1.4	745	16	US-10-140-864-68	Sequence 68, App1
44	40.8	1.4	255	13	US-10-085-783A-34088	Sequence 34088, A
45	40.8	1.4	255	16	US-10-242-535A-34088	Sequence 34088, A

ALIGNMENTS

RESULT 1
US-09-925-301-327
Sequence 327, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05892
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 327
LENGTH: 1764
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1758)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1759)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1762)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-327
Query Match 19.5%; Score 558; DB 9; Length 1764;
Best Local Similarity 74.2%; Pred. No. 3,4e-166;

OY 821 CTGGAG 827
DB 45 CTGTGG 39

RESULT 3

US-09-918-995-19127
Sequence 19127, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:

APPLICANT: Hyeq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT FILING DATE: 2001-07-30
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 19127
LENGTH: 478
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: misc feature
LOCATION: (1)...(478)
OTHER INFORMATION: n = A,T,C or G

US-09-918-995-19127

Query Match 10.0%; Score 288; DB 10; Length 478;
Best Local Similarity 77.7%; Pred. No. 2e-80;

Matches 348; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

OY 850 CAGACAGACCTCTCTGTAAGCAGAGCTGAGCTGAGGACCACTGCTCCAGTCA 909
DB 31 CGGACAGACCTCTGTAAGCAGAGCTGAGGACCACTGAGGAGGCTGCTCCAGGCT 90

OY 910 GAGACAGATGACATGAAATCATCTGGAAGGAGGACCTGAGGAAATACAGAGAA 969
DB 91 GGAAGAGATGACAGAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 150

OY 970 AACCCAGGAGAGCGGCAAAACCAACCTGAGTGTAGCTTACCAACCCCAATGCCCTG 1029
DB 151 GAGCTGAGAGCTGTGGCCGACGAGGAGCTGATGTAGAGGTTACACCCCAAGGCCCTG 210

OY 1030 TACTGATCTGCGCGCAGCTTCAACAACAAGTTATGATCTGTGATGCTGTGAG 1089
DB 211 TATTGATTTGCGCGCAGCTTCAACAACAAGTTATGATCTGTGATGCTGTGAG 270

OY 1090 GAGTGTTCATGATGATCTGTGAGGCTTCTGAGGCGCGAGGCGGCTCCCTGAGAA 1149
DB 271 GATGATTTTCAATGATGATCTGTGAGGCTTCTGAGGCGCGAGGCGGCTTCTGAGAA 330

OY 1150 AACGGGAGAGATACATCTGCCCAATTGCAACCATTTTGCAGATGAGATGAGACAAAC 1209
DB 331 AATGGGAGAGATATATCTGCCCAATGCAACCATTTTGCAGATGAGATGAGACATCAT 390

OY 1210 GGTAGCGCACCAATGAGAGAGCTGTGGGTGAGATCTGTGGGTCTGTGAGCAACAC 1269
DB 391 TCAGAAACCGCAGATCAGCAGAGAGAACTAAATGAGAGACCTGAGATGCTGATGCA 450

OY 1270 TGCAGAGATAGGAGCAGATGAGCAGA 1297
DB 451 TGTACAGATATAGAAACAATAGACAGA 478

RESULT 4

US-10-029-386-11381/C
Sequence 11381, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE

FILE REFERENCE: ABOMICA-X-2
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11381
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: MAP TO AL117379.14
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: SWISSPROT HIT: P80398, EVALUATE 3.80e+00
OTHER INFORMATION: NT HIT: g114769835, EVALUATE 1.00e-78
OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALUATE 2.00e-78

US-10-029-386-11381

Query Match 4.1%; Score 117.8; DB 15; Length 536;
Best Local Similarity 83.2%; Pred. No. 4.4e-26;

Matches 134; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 211 GCTTCAAGTTTGGACAGATATGATGATTAAGGACCTGAGCAATGAGAGACACC 270
DB 167 GCTTCAAGTTTGGACAGATATGATGATTAAGGACCTGAGCAATGAGAGACACC 108

OY 271 AAGCTATCAAAACCCAGGTAAGAGTTTCAAGAAAACCTGGGTTTTCAGAAACCAACG 330
DB 107 AAGGCTATCAAAACCCAGGTAAGAGTTTCAAGAAAACCTGGGTTTTCAGAAACCAACG 48

OY 331 ATTGCCAAGCTGAGGCTGAGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 371
DB 47 ATGCCAAGCTGAGGCTGAGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7

RESULT 5

US-09-908-975-25239
Sequence 25239, Application US/09908975
Publication No. US20030165843A1
GENERAL INFORMATION:

APPLICANT: WASSERMAN, Alon
APPLICANT: MINTZ, Eli
APPLICANT: PALMER, Simon
TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
FILE REFERENCE: 36688-0005
CURRENT FILING DATE: 2001-07-20
PRIOR FILING DATE: 2001-05-02
PRIOR APPLICATION NUMBER: US 60/287,724
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 32337
SOFTWARE: PatentIn version 3.0
SEQ ID NO 25239
LENGTH: 65
TYPE: DNA
ORGANISM: Mus musculus

US-09-908-975-25239

Query Match 2.3%; Score 65; DB 10; Length 65;
Best Local Similarity 100.0%; Pred. No. 6.9e-10;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2331 TTGAAGGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2390
DB 1 TTGAAGGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60

QY 2391 AGCAG 2395
Db 61 AGCAG 65

RESULT 6

US-09-764-864-471
Sequence 471, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 471
LENGTH: 915
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (894)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (903)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (905)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (914)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-471

Query Match 2.0%; Score 57.8; DB 9; Length 915;
Best Local Similarity 66.4%; Pred. No. 9.2e-07;

Matches 83; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1001 AGTGTAGATTACAGCCCATGCTGTACTGATCGCCGACGCTCACAACA 1060
Db 281 AGTGTAGATTACAGCCCATGCTGTACTGATCGCCGACGCTCACAACA 340
QY 1061 GGTATTAGATCTGCTGTATGCTGTAGAGAGTGTTCAGTGTGTGGTATT 1120
Db 341 GGTTCATATCGGGGTGTGCAACTGCATAGAGTGTTCATGGGAGCTGCATCCGATCA 400
QY 1121 CTGAG 1125
Db 401 CTGAG 405

RESULT 7

US-10-175-523-142
Sequence 142, Application US/10175523
Publication No. US20030096264A1
GENERAL INFORMATION:
APPLICANT: Brockman, Jeffrey
APPLICANT: Evans, David
APPLICANT: Hook, Derek
APPLICANT: Klimczak, Leszek
APPLICANT: Laeng, Pascal
APPLICANT: Paley, Michael
APPLICANT: Rajan, Pithi
TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPRTS)
FILE REFERENCE: 3235/13795-US3
CURRENT APPLICATION NUMBER: US/10/175,523
CURRENT FILING DATE: 2002-06-18
Prior application number: US 60/299,151
Prior filing date: 2001-06-18
Prior application number: US 60/317,828

PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 60/349,936
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/361,834
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn version 3.1
SEQ ID NO 142
LENGTH: 2320
TYPE: DNA
ORGANISM: Homo sapiens
US-10-175-523-142

Query Match 2.0%; Score 57.8; DB 15; Length 2320;
Best Local Similarity 66.4%; Pred. No. 1.8e-06;

Matches 83; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1001 AGTGTAGATTACAGCCCATGCTGTACTGATCGCCGACGCTCACAACA 1060
Db 109 AGTGTAGATTACAGCCCATGCTGTACTGATCGCCGACGCTCACAACA 168
QY 1061 GGTATTAGATCTGCTGTATGCTGTAGAGAGTGTTCAGTGTGTGGTATT 1120
Db 169 GGTTCATATCGGGGTGTGCAACTGCATAGAGTGTTCATGGGAGCTGCATCCGATCA 228
QY 1121 CTGAG 1125
Db 229 CTGAG 233

RESULT 8

US-09-873-367C-112
Sequence 112, Application US/09873367C
Publication No. US20030165839A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
APPLICANT: Soppet, Daniel
APPLICANT: Endress, Gregory
APPLICANT: Augustus, Meena
APPLICANT: Ebner, Reinhard
APPLICANT: Carter, Kenneth
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
FILE REFERENCE: 689290-64
CURRENT APPLICATION NUMBER: US/09/873,367C
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: U.S. 60/236,891
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/236,842
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/244,867
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: U.S. 60/245,084
PRIOR FILING DATE: 2000-11-01
NUMBER OF SEQ ID NOS: 1067
SOFTWARE: PatentIn version 3.0
SEQ ID NO 112
LENGTH: 5975
TYPE: DNA
ORGANISM: Homo sapiens
US-09-873-367C-112

Query Match 1.9%; Score 55.4; DB 10; Length 5975;
Best Local Similarity 59.0%; Pred. No. 2e-05;

Matches 95; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1018 CCCAATGCCCTGATCTGCTGCGCCGACGCTCACAACAAGTTATGATGCTCT 1077
Db 1200 CCCAATGCCCTGATCTGCTGCGCCGACGCTCACAACAAGTTATGATGCTCT 1259

Qy	1078	GATCGGTTGACGAGTGGTC	CATGGGACATGTGGGATATTTCTGAGGCCGAGGGCG	1137
Db	1260	GGAGAGATGTATGACTGCTGTTTCATGGTGATTTGTGTGGGTTAAAGTCCTTCTCAAGCACAG	1319	
Qy	1138	CTCTTGAAAGAAACGGGGAAGACTATCATCTGCCCAAATTG	1178	
Db	1320	CAGATGGCCGAGGAGACAAAGAAATATGTCTGTGTAATAAT	1360	

RESULT 9
HE-10-017-270-269

Sequence 269, Application US/10037270
Publication No. US20030104529A1

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Duntui
APPLICANT: Wang, Zhilwei
APPLICANT: Tillinghast, John
APPLICANT: Drmanac, Radje T.

TITLE OF INVENTION: No. US20030104529A1 Nucleic Acids and
TITLE OF INVENTION: Polypeptides

FILE REFERENCE: 784CIP2B

CURRENT APPLICATION NUMBER: US/10/037,270

CURRENT FILING DATE: 2002-01-04

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/488,725

PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1104

SOFTWARE: pt_pl_genes Version 1.0

SEQ ID NO 269

LENGTH: 6256

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: CDS
LOCATION: (257)..(5443)
US-10-037-270-269

Query Match	Similarity	1.9%	Score 55.4	DB 15	Length 6256
Best Local	Similarity	59.0%	Pred. No. 2.1e-05		
Matches	95	Conservative	0	Mismatches	66
				Indels	0
				Gaps	0
Qy	1018	CCCAATGCCCTGTACTGATCTGCCGCCGCTCACAACAACAGGTTTATGATCTGCTGT			1077
Db	1469	CCCAAGCAAGAGATGTGGGTTTGTGCAAAAACCAATGCCAACAGTTTATGTTGGCTGT			1528
Qy	1078	GATGCGTGTGAGAGATGTTCCATGCTGACCTGTGTGGTATTTCTGAGGCCCCGAGGGGGG			1137
Db	1529	GAGAGATGTGTGCTGGTTTCATGGTGATGTGTGTGGGTTTAAGTCCTTTCACAGCAG			1588
Qy	1138	CTCCGTGAAAGGAACGGGGAAGACTATCATCTGCCCAATTG			1178
Db	1589	CAGATGGCCGAGAGAACAAAGATATATGTCGTGTAAATG			1629

RESULT 10

US-10-117-722-269
; Sequence 269, Application US/10117722
; Publication No. US20030219744A1

```

GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Drianac, Radoje T.
TITLE OF INVENTION: No. US2003021974A1el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2BCIP
CURRENT APPLICATION NUMBER: US/10/117,722
CURRENT FILING DATE: 2002-04-04
PRIOR APPLICATION NUMBER: 09/620,312
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pt_fl_genes Version 1.0
SEQ ID NO 269
LENGTH: 6256
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: CDS
LOCATION: (257)..(5443)
US-10-117-722-269

```

	Query Match	Similarity	1.9%	Score 55.4	DB 16	Length 6256
	Best Local	Similarity	59.0%	Pred. No. 2.1e-05		
	Matches	95	Conservative	0	Mismatches	66
					Indels	0
					Gaps	0
Qy	1018	CCCAATCCCTGTACTGATCTGCGCGAGCTTCACAACAAGATTATGATCTGCTGT				1077
Db	1469	CCCAAGCAAGCAGTGTGGGTTTGGCAAAAACACATGSCACAAGTTTATGTTGGCTGT				1528
Qy	1078	GATCGTGTGAGAGATGCTTCATGTCATGTCGTGTGGTATTTCTGAGAGCCCGAGGCGG				1137
Db	1529	GGGAGATGTGATGACTGCTGTTCAAGTGTGATGTGTGGTTTAACTCTTTCAGACACAG				1588
Qy	1138	CTCCTGGAAGGAAACGGGGGAAGTACATCATCTGCCCAATTG				1178
Db	1589	CAGATGGCCAGAGAAACAAAGATATATGTCGTGTAAATG				1629

RESULT 11

```

US-09-930-213-315
Sequence 315, Application US/09930213
Publication No. US20030170625A1
GENERAL INFORMATION:
APPLICANT: ROSENTHAL, ANDRE
APPLICANT: HINZMANN, BERND
APPLICANT: SCHAFER, REINHARD
APPLICANT: ZUBER, JOHANNES
APPLICANT: TCHER-NITSE, OLEG
APPLICANT: GRIPS, MARTIN
APPLICANT: HELINGGEL, MARTIN
APPLICANT: SCHEWITZ, ANNE-CHANTAL
APPLICANT: SEES, CHRISTINE
TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
FILE REFERENCE: ALABR-14
CURRENT APPLICATION NUMBER: US/09/930,213
CURRENT FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: DE 10004102.7
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 885
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 315
LENGTH: 6948
TYPE: DNA
ORGANISM: Homo sapiens
US-09-930-213-315

```


Db 242 NSLSYENWQSSIGFVNPVVEFLLEKIDESNPPLMRGLQSKNPSBEDPAAALRK 301
Qy 918 TGACATAGAAAATCAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 977
Db 302 QSASWALKCVKIGVDFVGRHVDAMNENKALIEDKQVVALVARGALVATKGSINRAI 361
Qy 978 GGAGGCGGCAACCAAGCCGTGAGTGTGAGTTTACAGCCCAATGCCCTGTACTGAT 1037
Db 362 EDFBLALNCPFHRRARXKLCCTLIVERGQGLEBEBEFLNABEYKKALALDTEPDAEBA 421
Qy 1038 CTGCGCCGACCTCAACAAGGTTTATGATCTGCTGTGATCGGTGAGAGAGTGT 1097
Db 422 LQKHKYQKSLERKQAEKKEKOKTKITSAEKLRLLEKRLKRRKRSSTSSSV 481
Qy 1098 CCATGCTACTGTGAGTATTTCTAGAGCCGAGGAGGAGGAGGAGGAGGAGGAG 1157
Db 482 SSADSVSSSSSSSGHGRHKKRNRBSRSSRSRASSNQIDNRDECPVPA 541
Qy 1158 AGACTACATCTGCCCAATTTGACCATTTTGCAGAGTGACAGATAGACAAAGGTAAGCC 1217
Db 542 NTSASFVNHKQVVEKLDKQDRLOVEKTOIKEDRCPLSSSLEIPDDPGVTSYLFKKLT 601
Qy 1218 CACCAATAGCAGAGCTCTGGGTGAGATCTGTGGTGTCTGATGCA 1264
Db 602 IKQPGAPSGDIPBEGVITIDDSIHVTPEDLQVGDMEVEDSGID 648

RESULT 14
US-10-085-783A-56748
; Sequence 56748, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: liev, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)..(5)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (30)..(30)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-56748

Query Match 1.7%; Score 48; DB 13; Length 469;
Best Local Similarity 62.5%; Pred. No. 0.00076;
Matches 75; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1059 CAGGTTTATGATCTGTGATGATGATGATGATGATGATGATGATGATGAT 1118
Db 116 CAGGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 175
Qy 1119 TTCTGAGGCGGAGGCGGCTCTGGAAGAAAGGAGGAGGAGGAGGAGGAGGAG 1178
Db 176 AAGTCTTCTCAAGCACAGCAGATGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 235

RESULT 15
US-10-242-535A-56748
; Sequence 56748, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: liev, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)..(5)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (30)..(30)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-56748

Query Match 1.7%; Score 48; DB 16; Length 469;
Best Local Similarity 62.5%; Pred. No. 0.00076;
Matches 75; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Search completed: April 23, 2004, 23:50:40
Job time: 1366.62 secs

Qy 1059 CAGGTTTATGATCTGTGATGATGATGATGATGATGATGATGATGATGAT 1118
Db 116 CAGGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 175
Qy 1119 TTCTGAGGCGGAGGCGGCTCTGGAAGAAAGGAGGAGGAGGAGGAGGAGGAG 1178
Db 176 AAGTCTTCTCAAGCACAGCAGATGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 235

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:38:03 ; Search time 961.654 Seconds
(without alignment)
11529.921 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 2610
Sequence: 1 ctgcgtgcccgcgcgcac.....actcttaagatcatatcctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapept 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2610	100.0	2610	AA290578	AA290578 Human dea
2	2516.2	96.4	2772	AAH15102	AAH15102 Human CDN
3	1810.4	69.4	2085	ABV28029	ABV28029 Human pro
4	1810.4	69.4	2085	ABV27935	ABV27935 Human pro
5	1810.4	69.4	2085	ABV22095	ABV22095 Human pro
6	1808.8	69.3	2085	ABV22192	ABV22192 Human pro
7	1564.4	59.9	1764	AACT7933	AACT7933 Human cen
8	1280.6	49.1	7290	AAK53352	AAK53352 Human pol
9	1280.6	49.1	7838	AAK52368	AAK52368 Human pol
10	1190.8	45.6	2332	AAK86576	AAK86576 DNA encod
11	1023.6	39.2	2867	AA290579	AA290579 Murine de
12	948.4	36.3	3062	AAK67051	AAK67051 Human imm
13	682.6	26.2	2986	AAK86575	AAK86575 DNA encod
14	596.8	22.9	659	AAK08973	AAK08973 Human bec
15	510.8	19.6	681	AAH08005	AAH08005 Human CDN
16	503	19.3	503	AAK86574	AAK86574 DNA encod
17	435	16.7	478	ACG31915	ACG31915 Human end
18	414.4	15.9	481	AAV42461	AAV42461 Human pro
19	336.6	12.9	542	AAH12970	AAH12970 Human CDN
20	314.2	12.0	362	ABV12528	ABV12528 Human pro
21	305.6	11.7	365	ABV42575	ABV42575 Human pro
22	305.6	11.7	365	ABV33664	ABV33664 Human pro
23	249.8	9.6	256	ABV03359	ABV03359 Human pro

24	235.2	9.0	408	AAA43780	AAA43780 Mouse sec
25	231.2	8.9	324	ABN21248	ABN21248 Human ORF
26	209	8.0	7838	AAK52368	AAK52368 Human pol
27	202.4	7.8	2210	AACT74383	AACT74383 Human sec
28	182.6	7.0	7174	ADC37602	ADC37602 Human nuc
29	137.2	5.3	636	ABV03329	ABV03329 Human pro
30	102.6	3.9	593	ABV12398	ABV12398 Human pro
31	101.2	3.9	179	ABV33540	ABV33540 Human pro
32	91.2	3.5	226	ABK08454	ABK08454 Human leu
33	78.2	3.0	6051	ABR29341	ABR29341 Drosophila
34	68.6	2.6	8751	ABR29340	ABR29340 Drosophila
35	65	2.5	5975	ABR161775	ABR161775 Colom ade
36	65	2.5	6256	AAI58390	AAI58390 Human pol
37	65	2.5	6256	ADB48359	ADB48359 Novel hum
38	65	2.5	6584	AAI60176	AAI60176 Human pol
39	65	2.5	6547	ACF34533	ACF34533 Gene enco
40	65	2.5	6948	AAH81806	AAH81806 Human dif
41	65	2.5	6996	AAK53025	AAK53025 Human pol
42	65	2.5	7354	AAK52041	AAK52041 Human pol
43	60	2.3	60	ABN33521	ABN33521 Human spl
44	57.4	2.2	2000	ADR71938	ADR71938 Rice gene
45	56.8	2.2	10732	AAA10594	AAA10594 Gene enco

ALIGNMENTS

RESULT 1	AA290578	standard; cDNA; 2610 BP.
ID	AA290578	
XX	AA290578;	
AC	AA290578;	
XX		
DT	19-JUN-2000	(first entry)
XX		
DB	Human death inducer-oblierator 1 (DIO-1) polypeptide encoding cDNA.	
XX		
KW	Death inducer-oblierator 1; DIO-1; cell death; cancer; tumour; human; autoimmune disease; cytostatic; immunosuppressive; antidiabetic; antirheumatic; antiinflammatory; antiproliferative; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	265..1953
FT		/*tag= a
FT		/product= "DIO-1"
XX		
PN	WO200015787-A1.	
XX		
PD	23-MAR-2000.	
XX		
PF	10-SEP-1999;	99WO-GB003019.
XX		
PR	10-SEP-1998;	98SE-00003069.
XX		
PR	17-SEP-1998;	98US-0100873P.
XX		
PA	(CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.	
XX		
PA	(BANN/) BANNERMAN D G.	
XX		
PI	Alonso CM, Domingo DG, Grandien A, Leonardo B, Martinez P;	
XX		
DR	WPI: 2000-271426/23.	
XX		
DR	P-PSDB; AAY67579.	
XX		
PT	New DNA encoding human and murine death inducer-oblierator 1	
XX	polypeptides, useful in the treatment of cancer, autoimmune diseases,	
PT	diabetes, rheumatoid arthritis, benign tumors, malignant tumors and	
XX	hyperproliferative skin disorders.	
XX		
PS	Claim 1; Fig 1A; 27pp; English.	
XX		
CC	The invention provides nucleic acids encoding the human and murine death	

inducer-obliiterator 1 (DIO-1) polypeptides. The polypeptides can be expressed by standard recombinant methodology. The DIO-1 polypeptides, agonists and antagonists are used as a medicament for treating diseases characterized by an alteration in cell death or by hyperproliferation, e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumours, malignant tumours or hyperproliferative skin disorders. They are also useful in the treatment of metabolic, proliferative or inflammatory conditions. The present sequence represents a cDNA encoding the human DIO-1 polypeptide

Sequence 2610 BP; 644 A; 667 C; 734 G; 565 T; 0 U; 0 Other;

Query Match 100.0%; Score 2610; DB 3; Length 2610;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTCGGTGGCCGCTCCGCTCACTCCGCGGCTTCGGGGAAATGCTGCGAGACCTTAGAGGC 60
DB 1 CTCGGTGGCCGCTCCGCTCACTCCGCGGCTTCGGGGAAATGCTGCGAGACCTTAGAGGC 60
QY 61 CTGGGAGCTTACTCCACGGGACAGCTCTAGATAATCTGAGTTGTGAATAATACAGAG 120
DB 61 CTGGGAGCTTACTCCACGGGACAGCTCTAGATAATCTGAGTTGTGAATAATACAGAG 120
QY 121 CTTGTACTCGTGAACAGTGGCTGACAAACAGTGTGTGTGAGCCTGGCTGTCTGTGG 180
DB 121 CTTGTACTCGTGAACAGTGGCTGACAAACAGTGTGTGTGAGCCTGGCTGTCTGTGG 180
QY 181 ACCGAGAGTTTCGTCTCCAGAGGTTTTGTGTGATTTAGATTTCAGGAAAAGTGTCT 240
DB 181 ACCGAGAGTTTCGTCTCCAGAGGTTTTGTGTGATTTAGATTTCAGGAAAAGTGTCT 240
QY 241 CAAGCTTTCAGTGTGAGAGAGTATGAGACGACAAAGGCGACCCGAGACATGAGAGCA 300
DB 241 CAAGCTTTCAGTGTGAGAGAGTATGAGACGACAAAGGCGACCCGAGAGCA 300
QY 301 CCTAAGCCATCAACACCAACGACAAAGATTTCAGAAAATCATGGGGTTTTGAAAGACC 360
DB 301 CCTAAGCCATCAACACCAACGACAAAGATTTCAGAAAATCATGGGGTTTTGAAAGACC 360
QY 361 ACTATCCGCAAGCGAGAGGCGCAGGGGACCGAGAGCTGACCCACTGAGACCCGACCC 420
DB 361 ACTATCCGCAAGCGAGAGGCGCAGGGGACCGAGAGCTGACCCACTGAGACCCGACCC 420
QY 421 CCAAGAGAGAGCTGGGCTGTCTCCCTGGGCGCAGTGGAGAGCGCCAACTGCACTAG 480
DB 421 CCAAGAGAGAGCTGGGCTGTCTCCCTGGGCGCAGTGGAGAGCGCCAACTGCACTAG 480
QY 481 CGCGTGAAGAGCTTCTGACCAATTTGGGGGCGCCGCGCAGAGAGAGCATTCCTGTCTCC 540
DB 481 CGCGTGAAGAGCTTCTGACCAATTTGGGGGCGCCGCGCAGAGAGAGCATTCCTGTCTCC 540
QY 541 CTGAGAGATTCTGTGAGACCCAGCTCTGCCCCGACAGACGCGGAGACAGCTCCGAG 600
DB 541 CTGAGAGATTCTGTGAGACCCAGCTCTGCCCCGACAGACGCGGAGACAGCTCCGAG 600
QY 601 GCGAGCGTGAAGACGCTTCTGAGACGAGAGCGGCGCCGAGCTGTCTTCCACAGCTGTG 660
DB 601 GCGAGCGTGAAGACGCTTCTGAGACGAGAGCGGCGCCGAGCTGTCTTCCACAGCTGTG 660
QY 661 AAGGAGACGACGAGCTCTTCTGAAAAGGTGAAAAGGAGGAGTGAACAGATGACACTCC 720
DB 661 AAGGAGACGACGAGCTCTTCTGAAAAGGTGAAAAGGAGGAGTGAACAGATGACACTCC 720
QY 721 GATATGTCACAGATGAGCTTGAAGAGCTTGAAGAGCTTCAAGATTCGCTTGCAGAGGCG 780
DB 721 GATATGTCACAGATGAGCTTGAAGAGCTTGAAGAGCTTCAAGATTCGCTTGCAGAGGCG 780
QY 781 GAAAGAGGACCCACTGAGAGGCGCCCTGAAAAGGATCCAGAGTGGCTGCGGAAAGAGGCG 840
DB 781 GAAAGAGGACCCACTGAGAGGCGCCCTGAAAAGGATCCAGAGTGGCTGCGGAAAGAGGCG 840
QY 841 CGGAGAGAGGCTCCGCGGAGACTGTGGCTCCGAGGCGAGTGAACCTGTGAGAGGCGCTC 900
```

```
DB 841 CGGAGAGAGGCTCCGCGGAGACTGTGGCTCCGAGGCGAGTGAACCTGTGAGAGGCGCTC 900
QY 901 CTGCCCGATTAGCAGAGACCCGAGACGATCAGGGGTTGTGTCCCAAGCTGGGAAAAT 960
DB 901 CTGCCCGATTAGCAGAGACCCGAGACGATCAGGGGTTGTGTCCCAAGCTGGGAAAAT 960
QY 961 GACAGAGAGATGATTGAGAGGAGGAGGCTCAGGACATCAAGATGAGAGGAGCTGGA 1020
DB 961 GACAGAGAGATGATTGAGAGGAGGAGGCTCAGGACATCAAGATGAGAGGAGCTGGA 1020
QY 1021 GACTTGGCCGACCGAAGCTGAATGTGAGGTTACGACCCCAAGCCCTGTATTGCAAT 1080
DB 1021 GACTTGGCCGACCGAAGCTGAATGTGAGGTTACGACCCCAAGCCCTGTATTGCAAT 1080
QY 1081 TGCAGCCAGCTTCAACACAGAGTTTATGATTTGTCTGTACCCGTGTGAAGAAATGTTT 1140
DB 1081 TGCAGCCAGCTTCAACACAGAGTTTATGATTTGTCTGTACCCGTGTGAAGAAATGTTT 1140
QY 1141 CATGGCAGATTGTGAGGCAATTTCTGAGGCTGAGGAGGCTTTTGAAGAAATGGGAA 1200
DB 1141 CATGGCAGATTGTGAGGCAATTTCTGAGGCTGAGGAGGCTTTTGAAGAAATGGGAA 1200
QY 1201 GACTATATCTGCCCCAACTGACCAATTCGCAAGTGCAGATGAGACTCAATTCAGAAA 1260
DB 1201 GACTATATCTGCCCCAACTGACCAATTCGCAAGTGCAGATGAGACTCAATTCAGAAA 1260
QY 1261 GCAGATCAGCAGAAAGCTAATATGAGACCTGAGATGCTGATGCGCACTGATTCAGAT 1320
DB 1261 GCAGATCAGCAGAAAGCTAATATGAGACCTGAGATGCTGATGCGCACTGATTCAGAT 1320
QY 1361 GCAGATCAGCAGAAAGCTAATATGAGACCTGAGATGCTGATGCGCACTGATTCAGAT 1320
DB 1361 GCAGATCAGCAGAAAGCTAATATGAGACCTGAGATGCTGATGCGCACTGATTCAGAT 1320
QY 1321 ATAGGAACAATAGAGCAGAAAGCTTAGGCAAGACCAAGAGGATTAAGGATTAAGAAA 1380
DB 1321 ATAGGAACAATAGAGCAGAAAGCTTAGGCAAGACCAAGAGGATTAAGGATTAAGAAA 1380
QY 1381 GCTGCAAAATCCAGTGGCAGAAAGAACTCAAGATCTTCCAGCTGTGATGAGAGCGCT 1440
DB 1381 GCTGCAAAATCCAGTGGCAGAAAGAACTCAAGATCTTCCAGCTGTGATGAGAGCGCT 1440
QY 1441 GGTGCTCTCAAAATGATTTGGGCGCGGCTGTGTCACTGTGGGCGAGCCGACCTGGGTAC 1500
DB 1441 GGTGCTCTCAAAATGATTTGGGCGCGGCTGTGTCACTGTGGGCGAGCCGACCTGGGTAC 1500
QY 1501 TGCAATTAATGACTGTATCTCTCAAAACAGCGCCGAGGACAAATGAAGTTCTAAGCTCAG 1560
DB 1501 TGCAATTAATGACTGTATCTCTCAAAACAGCGCCGAGGACAAATGAAGTTCTAAGCTCAG 1560
QY 1561 AAGAGCAGAGCCAAAGCTTAAAGAAATGAATGAAGCCAGAGACCCAGCTT 1620
DB 1561 AAGAGCAGAGCCAAAGCTTAAAGAAATGAATGAAGCCAGAGACCCAGCTT 1620
QY 1621 CCGAATATCGGTGCTCAGGCGAGTATTAATCTCTTCTGTGACAAAGACAGCTCCA 1680
DB 1621 CCGAATATCGGTGCTCAGGCGAGTATTAATCTCTTCTGTGACAAAGACAGCTCCA 1680
QY 1681 GAAAAAAGAGAGCAACAGTGAAGAGGAGTGGTCCCTGCGCGGAGTGAAGCACTC 1740
DB 1681 GAAAAAAGAGAGCAACAGTGAAGAGGAGTGGTCCCTGCGCGGAGTGAAGCACTC 1740
QY 1741 GGGAGAGAGAGCAGCTTGTGAGAGCAGACGCGCTGTGGGCGAGGAGTCAAAATACAT 1800
DB 1741 GGGAGAGAGAGCAGCTTGTGAGAGCAGACGCGCTGTGGGCGAGGAGTCAAAATACAT 1800
QY 1801 GCAATTAAGCCAGAAAAGATCTGTCTCCCTCGCGCTCACTGTGTATTAATATGATAT 1860
DB 1801 GCAATTAAGCCAGAAAAGATCTGTCTCCCTCGCGCTCACTGTGTATTAATATGATAT 1860
QY 1861 CACCTAGGGGTTGGCTCTGAGACCCCTCCGTTCTTCTGAGATAGCCATCCCTGGGCG 1920
DB 1861 CACCTAGGGGTTGGCTCTGAGACCCCTCCGTTCTTCTGAGATAGCCATCCCTGGGCG 1920
QY 1921 TGTCCAGAGCTGGAGATTGAGCTTTGTGTAACTGATCAAGACCCGCTGCACTAT 1980
```



```

Db      1921 TGTCCAGGACTGGAGTTGCACTTGTGTGTTAAGTGAATCAGACACCGCGTCGACCAT 1980
Oy      1981 CAGCGGGAAGACGAGACCCCATGTCAGAGATGCTCTCGTCGCGCTGTGTGCATATCCCAAGTC 2040
Db      1981 CAGCGGGAAGACGAGACCCCATGTCAGAGATGCTCTCGTCGCGCTGTGTGCATATCCCAAGTC 2040
Oy      2041 TGTGAGAGCTTCTGTGCACTGTTTTCGAAAGCTGTAAACCTGACGTGTAAGAGCTTCACTT 2100
Db      2041 TGTGAGAGCTTCTGTGCACTGTTTTCGAAAGCTGTAAACCTGACGTGTAAGAGCTTCACTT 2100
Oy      2101 TAAATGATGATTTCTTAAATCTGTGTTTTCAGTCTGAGGCTGTGTAAGTATTGTATCT 2160
Db      2101 TAAATGATGATTTCTTAAATCTGTGTTTTCAGTCTGAGGCTGTGTAAGTATTGTATCT 2160
Oy      2161 CTTCATCTCCAGTCTGATGATGACACACCTGCGCGGACCGGACATTCACCCCTGTCTG 2220
Db      2161 CTTCATCTCCAGTCTGATGATGACACACCTGCGCGGACCGGACATTCACCCCTGTCTG 2220
Oy      2221 CACATGAGTGTCTGACAAACAGCGCTGTATAGCTTCAGTTTTCGACATTTGCCAGCG 2280
Db      2221 CACATGAGTGTCTGACAAACAGCGCTGTATAGCTTCAGTTTTCGACATTTGCCAGCG 2280
Oy      2281 CCAGCAGATGAAGAATCATCTTTTATGTTGTGGAATCTTTGCAAGTTAGTGTG 2340
Db      2281 CCAGCAGATGAAGAATCATCTTTTATGTTGTGGAATCTTTGCAAGTTAGTGTG 2340
Oy      2341 CATCTGATTTTCAGGTGTATCATTTATTTTGACTGGGAGATPAGGGGATTTTTTTTTT 2400
Db      2341 CATCTGATTTTCAGGTGTATCATTTATTTTGACTGGGAGATPAGGGGATTTTTTTTTT 2400
Oy      2401 TCCATGTCGATTCACACGCTACACACCCACATGAAACATTCGAACTTCGAAAGGACAC 2460
Db      2401 TCCATGTCGATTCACACGCTACACACCCACATGAAACATTCGAACTTCGAAAGGACAC 2460
Oy      2461 ACTCTGCTTCATYAGGCCCCCAGCGTATGAGTTCACACTGAGAACACTGTCTGACCGC 2520
Db      2461 ACTCTGCTTCATYAGGCCCCCAGCGTATGAGTTCACACTGAGAACACTGTCTGACCGC 2520
Oy      2521 AGGAGCGGTGCTTGTGACCTTGTATTTTCAATGTGACCTGCTTCTGCGCTGCTCTT 2580
Db      2521 AGGAGCGGTGCTTGTGACCTTGTATTTTCAATGTGACCTGCTTCTGCGCTGCTCTT 2580
Oy      2581 GAATGTTTGAATCTTAAAGATCATATCTG 2610
Db      2581 GAATGTTTGAATCTTAAAGATCATATCTG 2610

```

RESULT 2
AAH15102
ID AAH15102 standard; cDNA; 2772 BP.

```

XX      AAH15102;
AC      26-JUN-2001 (first entry)
XX      Human cDNA sequence SEQ ID NO:13129.
XX      Human, primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX      Homo sapiens.
XX      BP1074617-A2.
XX      07-FEB-2001.
XX      28-JUL-2000; 2000EP-00116126.
XX      29-JUL-1999; 99JP-00248036.
XX      27-AUG-1999; 99JP-00300253.
XX      11-JAN-2000; 2000JP-00118776.
XX      02-MAY-2000; 2000JP-00183767.
XX      09-JUN-2000; 2000JP-00241899.
XX

```

```

PA      (HELI-) HELIX RES INST.
XX      Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI      Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX      WPI; 2001-318749/34.
XX      Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT      length cDNAs defined in the specification, and for the detection and/or
PT      diagnosis of the abnormality of the proteins encoded by the full-length
XX      cDNAs.
XX      Claim 8; SEQ ID NO 13129; 2537bp + Sequence Listing; English.
XX
CC      The present invention describes primer sets for synthesizing 5602 full-
CC      length cDNAs defined in the specification, where a primer set comprises:
CC      (a) an oligo-dT primer and an oligonucleotide complementary to the
CC      complementary strand of a polynucleotide which comprises one of the 5602
CC      nucleotide sequences defined in the specification, where the
CC      oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC      of an oligonucleotide comprising a sequence complementary to the
CC      complementary strand of a polynucleotide which comprises a 5'-end
CC      sequence and an oligonucleotide comprising a sequence complementary to a
CC      polynucleotide which comprises a 3'-end sequence, where the
CC      oligonucleotide comprises at least 15 nucleotides and the combination of
CC      the 5'-end sequence/3'-end sequence is selected from those defined in the
CC      specification. The primer sets can be used in antisense therapy and in
CC      gene therapy. The primers are useful for synthesizing polynucleotides,
CC      particularly full-length cDNAs. The primers are also useful for the
CC      detection and/or diagnosis of the abnormality of the proteins encoded by
CC      the full-length cDNAs. The primers allow obtaining of the full-length
CC      cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC      AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC      represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC      present invention
XX
XX      Sequence 2772 BP; 699 A; 691 C; 765 G; 617 T; 0 U; 0 Other;
XX
XX      Query Match      96.4%; Score 2516.2; DB 4; Length 2772;
XX      Best Local Similarity 99.8%; Pred. No. 0;
XX      Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;
XX
Oy      65 GAGCTTACTCCAGCGGGAACAGCTTATGATATATCTGAGTTGTGAAATACGAAGCTTG 124
Db      101 GAGCTTACTCCAGCGGGAACAGCTTATGATATATCTGAGTTGTGAAATACGAAGCTTG 160
Oy      125 TTAATCTGTAACAGTGGCTGACAAACAGTGTGTTGAGCTGCTGTCTGCTTGACCC 184
Db      161 TTAATCTGTAACAGTGGCTGACAAACAGTGTGTTGAGCTGCTGTCTGCTTGACCC 220
Oy      185 AAGAGTTTGTCTGCGACAGGTTTGTGTTGATTTAGATTCAGGGAAGAGTGTCCAG 244
Db      221 AAGAGTTTGTCTGCGACAGGTTTGTGTTGATTTAGATTTAGGGAAGAGTGTCCAG 280
Oy      245 CTTCAGTGTGAGAGAGTATGACACAAAGGCGACCGGACCAATGAGAGGCACTTA 304
Db      281 CTTCAGTGTGAGAGAGTATGACACAAAGGCGACCGGACCAATGAGAGGCACTTA 340
Oy      305 AGGCAATCAACCCACAGCAAAAGATTGAGAAAACATGGGGTTTTCGAAAGACCACTA 364
Db      341 AGGCAATCAACCCACAGCAAAAGATTGAGAAAACATGGGGTTTTCGAAAGACCACTA 400
Oy      365 TGGCCAAAGGAGAGGCGGAGGAGGAGGCGGAGGCTGACCACTGAGAGCGGACCCGAC 424
Db      401 TGGCCAAAGGAGAGGCGGAGGAGGAGGCGGAGGCTGACCACTGAGAGCGGACCCGAC 460
Oy      425 AGCAGAGCTGGGCTGTCTGCTGCGCGCAGTGGAGAGGACCCGCAAGCGCATGAGCGG 484
Db      461 AGCAGAGCTGGGCTGTCTGCTGCGCGCAGTGGAGAGGACCCGCAAGCGCATGAGCGG 520
Oy      485 TGGAGAGTTCTGATCACTTGGCGGCGCGCGGCGGCGGAGAGAGCATGCTGTCTCTCTGG 544

```

D 521 TGGAGCAGTTCTTGACCAATTGGGGGCGCGCGGAGGAGGACATGCTGTCTCCCTGG 580
Q 545 AGGATTCGTGTAGCCCACTGCTCCGCCCAACAGCGCCGAGACAGCTCTCCAGGGCA 604
D 581 AGGATTCGTGTAGCCCACTGCTCCGCCCAACAGCGCCGAGACAGCTCTCCAGGGCA 640
Q 605 GCGGTGAAGGCGCTTCTGAGACCAAGAGCGGCCCACTGCTCTCCAGCTGTAGAG 664
D 641 GCGGTGAAGGCGCTTCTGAGACCAAGAGCGGCCCACTGCTCTCCAGCTGTAGAG 700
Q 665 AACGACAGCTCTTCTGAAAAAGTGAAGAGGGATGACACGATGACACTTCCGATA 724
D 701 AACGACAGCTCTTCTGAAAAAGTGAAGAGGGATGACACGATGACACTTCCGATA 760
Q 725 GTGACAGGATGCGCTGACTTTGAAAGCTTCAGATTCGCTTTCGAGAGCGCGGAC 784
D 761 GTGACAGGATGCGCTGACTTTGAAAGCTTCAGATTCGCTTTCGAGAGCGCGGAC 820
Q 785 AGGAGCCCACTGAGAGGCGCCGTAAGGGATTCAGAGTCGCTCCGAGAGAGCGCGG 844
D 821 AGGAGCCCACTGAGAGGCGCCGTAAGGGATTCAGAGTCGCTCCGAGAGAGCGCGG 880
Q 845 AGGAGGCTCCGCGCGAGACTGTGGCTCCAGGCCAGTGAACACTGTGAGAGGCGCTTC 904
D 881 AGGAGGCTCCGCGCGAGACTGTGGCTCCAGGCCAGTGAACAACACTGTGAGAGGCGCTTC 940
Q 905 CCACTGACAGAGAGCCCGAGAACGATCAGGGGGTGTGTCCAGGCTGGAAAAATGACA 964
D 941 CCACTGACAGAGAGCCCGAGAACGATCAGGGGGTGTGTCCAGGCTGGAAAAATGACA 1000
Q 965 GAGGAGTAAGTTGAGAGGAAAGCGGCTCAGACATCAAGATGAGAGGCTGTGAGACT 1024
D 1001 GAGGAGTAAGTTGAGAGGAAAGCGGCTCAGACATCAAGATGAGAGGCTGTGAGACT 1060
Q 1025 TGGGCGGACCGAAGCTGATATGTAGGGTTACGACCCCAACGCTCTGTATTGCAATTTG 1084
D 1061 TGGGCGGACCGAAGCTGATATGTAGGGTTACGACCCCAACGCTCTGTATTGCAATTTG 1120
Q 1085 GCCAGCTCAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAAGATGTTTCTATG 1144
D 1121 GCCAGCTCAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAAGATGTTTCTATG 1180
Q 1145 GCGATGTGTGGGCTTTCTGAGGCTCGAGGGAGCTTTTGAAGGAATGAGGAAAGACT 1204
D 1181 GCGATGTGTGGGCTTTCTGAGGCTCGAGGGAGCTTTTGAAGGAATGAGGAAAGACT 1240
Q 1205 ATATCTGCCCAAACTGCAACAATTCGCAAGTGCAGATGAGACTATTCAGAAACGGCAG 1264
D 1241 ATATCTGCCCAAACTGCAACAATTCGCAAGTGCAGATGAGACTATTCAGAAACGGCAG 1300
Q 1265 ATCTAGCAGAAAGCTTAATGAGAGCTTGAGATGCTGATGAGCAGCAATTTGTAAGTATG 1324
D 1301 ATCTAGCAGAAAGCTTAATGAGAGCTTGAGATGCTGATGAGCAGCAATTTGTAAGTATG 1360
Q 1325 GAAACAATGAGAGAGTCTAGCGAAGCAAGGGATTAAGGGTGAATTTGAAAGCTG 1384
D 1361 GAAACAATGAGAGAGTCTAGCGAAGCAAGGGATTAAGGGTGAATTTGAAAGCTG 1420
Q 1385 CAAATCCAAAGTGCAGAAAGAACTCAAGATCTTCAGGCTGTGATGAGGCGCTGTGTG 1444
D 1421 CAAATCCAAAGTGCAGAAAGAACTCAAGATCTTCAGGCTGTGATGAGGCGCTGTGTG 1480
Q 1445 CCTCAAAATGTAATGGCGCGGGTGTCTGACGTGGCGCAGCCGACATCGGTGTACTGCA 1504
D 1481 CCTCAAAATGTAATGGCGCGGGTGTCTGACGTGGCGCAGCCGACATCGGTGTACTGCA 1540
Q 1505 GTAATGACTGTATCTCTCAAAACGCGCGCAGCAGATGAAAGTTTCTAAGCTCAGGTAAG 1564
D 1541 GTAATGACTGTATCTCTCAAAACGCGCGCAGCAGATGAAAGTTTCTAAGCTCAGGTAAG 1600
Q 1565 AACAGAGCCAAAGCTTAAGAAAAAGATGAAAGTAAAGCCAGAGAACCCAGTCTTCCGA 1624
D 1601 AACAGAGCCAAAGCTTAAGAAAAAGATGAAAGTAAAGCCAGAGAACCCAGTCTTCCGA 1660

Q 1625 AATGCGGTGCTCAGGACAGTATTAATAATCTCTTCTGTGACACAAGAGACCACTCCAGAAA 1684
D 1661 AATGCGGTGCTCAGGACAGTATTAATAATCTCTTCTGTGACACAAGAGACCACTCCAGAAA 1720
Q 1685 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCCGCGGAGTGAAGCACTCGGGA 1744
D 1721 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCCGCGGAGTGAAGCACTCGGGA 1780
Q 1745 AGAAGACAGCTTGTGAGAGAGCAGACGCGGTGTGTGGGCGAGATGACAAATTAAGTGA 1804
D 1781 AGAAGACAGCTTGTGAGAGAGCAGACGCGGTGTGTGGGCGAGATGACAAATTAAGTGA 1840
Q 1805 TAAAGCCAGAAAAAGCTGCTGCTCCCTGCGGCTCACTGTGTATTAATGATGATCAAC 1864
D 1841 TAAAGCCAGAAAAAGCTGCTGCTCCCTGCGGCTCACTGTGTATTAATGATGATCAAC 1900
Q 1865 TAGGGGTTGGCTCTGTGACCCCTCCGCTTCTTGTGATAGCCATCCCTGGGCTGTG 1924
D 1901 TAGGGGTTGGCTCTGTGACCCCTCCGCTTCTTGTGATAGCCATCCCTGGGCTGTG 1960
Q 1925 CAGACCTGGAGTTGCAAGCTTTGTGTAAAGCTGAATCAAGACACGCGTGAACATCAGC 1984
D 1961 CAGACCTGGAGTTGCAAGCTTTGTGTAAAGCTGAATCAAGACACGCGTGAACATCAGC 2020
Q 1985 GGAAGACAGAGCCCAATGTCAGAGATGCTCTGCTGCGCTGTGTCAATCCCTAGTCTGT 2044
D 2021 GGAAGACAGAGCCCAATGTCAGAGATGCTCTGCTGCGCTGTGTCAATCCCTAGTCTGT 2080
Q 2045 AGGACTTCTGTCACTGTTTTCAAAGCTGTAACCTGACGTGTAAGCTTCACTTAAT 2104
D 2081 AGGACTTCTGTCACTGTTTTCAAAGCTGTAACCTGACGTGTAAGCTTCACTTAAT 2140
Q 2105 GATTGATCTTAAATCTCTGTTTTCACCTGAGGCTCTGTAAGTATTTGATCTCTTC 2164
D 2141 GATTGATCTTAAATCTCTGTTTTCACCTGAGGCTCTGTAAGTATTTGATCTCTTC 2200
Q 2165 ATCCAGCTGATGATGATGACACACTGCGCGGACAGCCACATCCCTGTGTGACA 2224
D 2201 ATCCAGCTGATGATGATGACACACTGCGCGGACAGCCACATCCCTGTGTGACA 2260
Q 2225 TGAATGTTCTGACAAACGCGCTGTAAAGCTTCAAGTTTTCACATGTGCCAGGCCAG 2284
D 2261 TGAATGTTCTGACAAACGCGCTGTAAAGCTTCAAGTTTTCACATGTGCCAGGCCAG 2320
Q 2285 CACATGAAGACATCTCTTTTAAATGTGTGGGAATCTTGAAGTATGATGATC 2344
D 2321 CACATGAAGACATCTCTTTTAAATGTGTGGGAATCTTGAAGTATGATGATC 2380
Q 2345 TGAATTTGAGGTGATCAATTTTGAATTTGACTGGGACAGATGAGGGAATTTTTCCTCA 2404
D 2381 TGAATTTGAGGTGATCAATTTTGAATTTGACTGGGACAGATGAGGGAATTTTTCCTCA 2438
Q 2405 TGTCCGATTCACACCTTACACACCACTGAACACATTTGAACTTGAAG--CAACACT 2463
D 2439 TGTCCGATTCACACCTTACACACCACTGAACACATTTGAACTTGAAG--CAACACT 2498
Q 2464 CCTGCTTCAATAGGCCCAACGGAATGAGTTCAACCTAAGAACATGCTCTGACGCGAG 2523
D 2499 CCTGCTTCAATAGGCCCAACGGAATGAGTTCAACCTAAGAACATGCTCTGACGCGAG 2558
Q 2524 ACGGTGCTTGAACCTTGTGATTTCAACATGTAAGCTTCTGCTCCCTGCTCTTGA 2583
D 2559 ACGGTGCTTGAACCTTGTGATTTCAACATGTAAGCTTCTGCTCCCTGCTCTTGA 2618
Q 2584 TGTTTAAGCTTTAAGATCATATCTG 2610
D 2619 TGTTTAAGCTTTAAGATCATATCTG 2645

RESULT 3
ABV28029
ID ABV28029 standard; cDNA; 2085 BP.

XX AC ABV28029;
XX DT 16-SBP-2002 (first entry)
XX DB Human prostate expression marker cDNA 28020.
XX KM Human; prostate cancer; cytosolic; carcinogen; pharmacodynamic marker;
XX KM pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN W0200160860-A2.
XX PD 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US005171.
XX PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0213114P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JB;
XX DR WPI; 2001-662795/76.
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer.
XX PS Claim 1; Page 5788; 11750p; English.
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for: (a) assessing whether
XX CC a patient is afflicted with prostate cancer; (b) monitoring the
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX CC determining whether prostate cancer has metastasized in a patient; (h)
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient
XX CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX CC
XX CC Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;
XX CC
XX CC Query Match 69.4%; Score 1810.4; DB 5; Length 2085;
XX CC Best Local Similarity 98.8%; Pred. No. 0;
XX CC Matches 1960; Conservative 0; Mismatches 11; Indels 13; Gaps 13;
XX CC
QY 3 CGGTGCGCGTCCGACCTCCGCGCGTTCGCGGAATGCTGCGAGACCTAGAGGCT 62
DB 17 CGGTGCGCGTCCGACCTCCGCGCGTTCGCGGAATGCTGCGAGACCTAGAGGCT 76
QY 63 GCGGAGCTTACTCCAGCGGAAACGCTCTAGATTAATCTGAGTTTGTGAATAACAGGC 122
DB 77 GCGGAGCTTACTCCAGCGGAAACGCTCTAGATTAATCTGAGTTTGTGAATAACAGGC 135
QY 123 TGTATCTGTGAACAGTGTGCTGCAACAGTGTGTGTGAGGCTGCTGTGCTTGAGC 182
DB 136 TGTATCTGTGAACAGTGTGCTGCAACAGTGTGTGTGAGGCTGCTGTGCTTGAGC 195
QY 183 CCAAGAGTTTGTCTGCTGCAAGGTTTGTGTATTTAGATTTTCAAGGAAAGTGTCCA 242
DB 196 CCAAGAGTTTGTCTGCTGCAAGG-TTTTGTGTATTTAGATTTTCAAGGAAAGTGTCCA 254
QY 243 AGCTTTCAGTGTGAGCAGGTATGAGCAACAAAGCGACCCAGACATGAGAGGACCC 302

DB 255 AGCTTTCAGTGTGAGCAGGTATGAGCAACAAAGCGACCCAGACATGAGAGGACCC 313
QY 303 TAAAGCCATCAAAACCCACAGCAAAAGTTTCAGGAAACATAGGGTTTTCGAGAGCAGC 362
DB 314 TAAAGCCATCAAAACCCACAGCAAAAGTTTCAGGAAACATAGGGTTTTCGAGAGCAGC 373
QY 363 TATGCGCAAGGAGAGGCGCAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 422
DB 374 TATGCGCAAGGAGAGGCGCAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 432
QY 423 ACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 482
DB 433 ACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 492
QY 483 CGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 542
DB 493 CGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 552
QY 543 GAGAGAGTCTGAGTGAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 602
DB 553 GAGAGAGTCTGAGTGAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 610
QY 603 CAGCGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 662
DB 611 CAGCGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 668
QY 663 GGAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 722
DB 669 GGAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 727
QY 723 TAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 782
DB 728 TAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 785
QY 783 ACAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 842
DB 786 ACAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 845
QY 843 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 902
DB 846 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 905
QY 903 GCCAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 962
DB 906 GCCAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 965
QY 963 CAGAGAGGAGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1022
DB 966 CAGAGAGGAGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1025
QY 1023 CTTGGGCGCAGCGAAGGCTGAATGTAGGGTTTACGACCCCAAGCGCCTGTATTCATTGG 1082
DB 1026 CTTGGGCGCAGCGAAGGCTGAATGTAGGGTTTACGACCCCAAGCGCCTGTATTCATTGG 1085
QY 1083 CCGCAGGCTCAACAACAAGGTTTATGATTTGCTGTGACCCCTGTGAAGAAATGCTTCA 1142
DB 1086 CCGCAGGCTCAACAACAAGGTTTATGATTTGCTGTGACCCCTGTGAAGAAATGCTTCA 1145
QY 1143 TGGCGATTTGTGGGCAATTTCTGAGGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1202
DB 1146 TGGCGATTTGTGGGCAATTTCTGAGGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1205
QY 1203 CTATATCTGCCCAACTGACCAATTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1262
DB 1206 CTATATCTGCCCAACTGACCAATTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1265
QY 1263 AGATGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1322
DB 1266 AGATGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1325
QY 1323 AGGAACATTAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1382

603 CAGCGTGAAGACGCTTCTGAGACCAAGACGCGCCCGCAGTCTGCTTCCACAGCTGTGA 662
 611 CAGGCTGGAAGACGCTTCTGAGACCAAGACGCG -CCCCAGTCTGCTTCCACAG -TGTGA 668
 663 GGAACGACAGCTTCTTCTGAAAAAGTGAAGAGGGGATGACCAAGATGACCACTCCGA 722
 669 GGAACGACAG -CTTCTTCTGAAAAAGTGAAGAGGGGATGACCAAGATGACCACTCCGA 727
 723 TAGTGAACAGGATGAGCTGACCTTGAAGAGCTTCAAGATGAGCTTCCAGAGAGCGGA 782
 728 TAGTGAACAGGATGAG -CTGACCTTGAAGAGCTTCAAGATGAG -CTTCCAGAGAGCGGA 785
 783 ACAGAGGCCCACTGAGAGAGCCCTGTAAGAGGATCCAGAGTGCCTGCGAGAGAGCGCG 842
 786 ACAGAGGCCCACTGAGAGAGCCCTGTAAGAGGATCCAGAGTGCCTGCGAGAGAGCGCG 845
 843 GAGAGAGGCTCCCGCGGAGACTGTGGGCTCCGAGGCCAGTGAACATGTGAGAGGGGCTCT 902
 846 GAGAGAGGCTCCCGCGGAGACTGTGGGCTCCGAGGCCAGTGAACATGTGAGAGGGGCTCT 905
 903 GCCCAGTAAGAGAGAGCCCGAGAGACATCAAGAGGGGTTGTCTCCAGAGCTGGAGAAAGTGA 962
 906 GCCCAGTAAGAGAGAGCCCGAGAGACATCAAGAGGGGTTGTCTCCAGAGCTGGAGAAAGTGA 965
 963 CAGAGAGATGAAGTTGAG 1022
 966 CAGAGAGATGAAGTTGAG 1025
 1023 CTTGGGCGGAG 1082
 1026 CTTGGGCGGAG 1085
 1083 CCGCAGAGCTTCAACAACAGAGTTATGATTTGCTGTGACCGCTGTGAAGAGAGAGTTTCA 1142
 1086 CCGCAGAGCTTCAACAACAGAGTTATGATTTGCTGTGACCGCTGTGAAGAGAGAGTTTCA 1145
 1143 TGGCGATTGTGGGCAATTTCTGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1202
 1146 TGGCGATTGTGGGCAATTTCTGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1205
 1203 CTAATCTGCGCCAACTGCAACAATCTGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1262
 1206 CTAATCTGCGCCAACTGCAACAATCTGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1265
 1263 AGATCAGCAGAGAGCTTAATGAG 1322
 1266 AGATCAGCAGAGAGCTTAATGAG 1325
 1323 AGGAACTAATAG 1382
 1326 AGGAACTAATAG 1385
 1383 TGCAGATCCAG 1442
 1386 TGCAGATCCAG 1445
 1443 TGCCTCAAAATGTAATGAG 1502
 1446 TGCCTCAAAATGTAATGAG 1505
 1503 CAGTAATGAG 1562
 1506 CAGTAATGAG 1565
 1563 AGAAG 1622
 1566 AGAAG 1625
 1623 GAAATGCGGTGCTCAGGAG 1682
 1626 GAAATGCGGTGCTCAGGAG 1685
 1683 AAAAAAG 1742

1686 AAAAAAG 1745
 1743 GAAAG 1802
 1746 GAAAG 1805
 1803 AGTAAG 1862
 1806 AGTAAG 1865
 1863 CTAAG 1922
 1866 CTAAG 1924
 1923 TCCAG 1982
 1925 TCCAG 1983
 1983 GCGG 1986
 1984 GCAG 1987
 1987 GCAG 1987
 RESULT 5
 ABV22095
 ID ABV22095 standard; cDNA; 2085 BP.
 XX
 AC ABV22095;
 DT 13-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 22086.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-021314P.
 PR 18-JUL-2000; 2000US-021907P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JB;
 XX
 DR MPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 3785; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Query Match 69.4%; Score 1810.4; DB 5; Length 2085;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1960; Conservative 0; Mismatches 11; Indels 13; Gaps 13;

```
QY 3 CGGTGGCCGTCCGCGCACTCCGCGCGCTTCGCGGAAATGCGTGCAGACCTTAGAGGCT 62
DB 17 CGGTGGCCGTCCGCGCACTCCGCGCGCTTCGCGGAAATGCGTGCAGACCTTAGAGGCT 76
QY 63 GCGGAGCTTACTCAGCGGAAACAGCTCTAGATAATCTGAATTGTTGAAATAGAAACC 122
DB 77 GCGGAGCTTACTCAGCGGAAACAGCTCTAGATAATCTGAATTGTTGAAATAGAAACC 135
QY 123 TGTACTGTGAAACAGTGGCTGACAGACAGTGTGTGTGAGCTGTGCTGTGAGAC 182
DB 136 TGTACTGTGAAACAGTGGCTGACAGACAGTGTGTGTGAGCTGTGCTGTGAGAC 195
QY 183 CAGAGGTTTCTGTGCGCAGGGTTTGTGTGATTTAGATTTCAAGGAAAGTGTCCA 242
DB 196 CAGAGGTTTCTGTGCGCAGGG-TTTTGTGTGATTTAGATTTCAAGGAAAGTGTCCA 254
QY 243 AGCTTTCAGTGTGAGACAGTATGACAGACAAAGCGACCCGAGCAATGAGAGGCAAC 302
DB 255 AGCTTTCAGTG-TGAGACAGGTATGACAGACAAAGCGACCCGAGCAATGAGAGGCAAC 313
QY 303 TAAAGGCATCAAAACCAACAGCAAAAGATTCAAGAAAACATGAGGTTTTCGAAAGACAC 362
DB 314 TAAAGGCATCAAAACCAACAGCAAAAGATTCAAGAAAACATGAGGTTTTCGAAAGACAC 373
QY 363 TATGCGCAAGGAGAGGCGCGAGGAGACGCGAGAGCTGACCTGAGGCCACACCC 422
DB 374 TATGCGCAAGGAGAGGCGCGAGGAGAGCGCGAGAGCTGACCTGAGGCCG-CACCC 432
QY 423 ACAGACAGCTGTGGGCTGTCTCTGCGCGCGAGTGGAGGAGGCCCAAGGCCACATGAG 482
DB 433 ACAGACAGCTGTGGGCTGTCTCTGCGCGCGAGTGGAGGAGGCCCAAGGCCACATGAG 492
QY 483 CGTGGAGCAGTTCTGACCAATTTGCGCGCGCGCGAGAGAGCATGCTGTCTCCCT 542
DB 493 CGTGGAGCAGTTCTGACCAATTTGCGCGCGCGCGCGAGAGAGCATGCTGTCTCCCT 552
QY 543 GAGAGATTCTGTGAGCCCAAGTCTCTGCGCGCGCAAGACGCGGAGACGCTTCGAGG 602
DB 553 GAGAGATTCTGTGAGCCCAAGTCTCTGCGCGCG-CACAGACGCGGAGACG-CTCCGAGG 610
QY 603 CAGCTGTGAAAGCGCTTCTGAGACCAAGAGGCGCCCAAGTCTCTTCCAGAGCTGTGAA 662
DB 611 CAGCTGTGAAAGCGCTTCTGAGACCAAGAGGCG-CGCCAGTCTGTCTCAAG-TGTGAA 668
QY 663 GGAACGACCAAGCTCTTCTGAAAAGTGAAGAGGAGATGACAGATGACACTCCGA 722
DB 669 GGAACGACCAAG-CTCTTCTGAAAAGTGAAGAGGAGATGACAGATGACACTCCGA 727
QY 723 TAGTGAACGAGTATGAGCTTGAAGAGCTTCAAGATGAGCTTTCGAGAGAGCGGAA 782
DB 728 TAGTGAACGAGTATGAG-CTGACCTTGAAGAGCTTCAAGATGAG-CTTTCGAGAGAGCGGAA 785
QY 783 ACAGAGACCCCACTGAGAGGCGCGTGAAGAGGATCCAGATGCGCTGCGGAAAGAGCGCG 842
DB 786 ACAGAGACCCCACTGAGAGGCGCGTGAAGAGGATCCAGATGCGCTGCGGAAAGAGCGCG 845
QY 843 GGAAGAGGCTCCCGCGAGACTGTGTGAGCTCCAGAGGCAAGTGAACATGTGAGAGGCTCT 902
DB 846 GGAAGAGGCTCCCGCGAGACTGTGTGAGCTCCAGAGGCAAGTGAACATGTGAGAGGCTCT 905
QY 903 GCCCAGTAAAGAGAGCCCGAGAAAGATCAAGAGGCTGTGTCCAGGCTGCGGAAAGATGA 962
DB 906 GCCCAGTAAAGAGAGCCCGAGAAAGATCAAGAGGCTGTGTCCAGGCTGCGGAAAGATGA 965
```

```
QY 963 CAGAGAGTATGTTGAGAGGAAAGCGGCTCAGAGACATCAAGATGAGAGGCTGAGGA 1022
DB 966 CAGAGAGTATGTTGAGAGGAAAGCGGCTCAGAGACATCAAGATGAGAGGCTGAGGA 1025
QY 1023 CTTGGGCTGACCGAAGCTTGAATGTGAGGTTACGACCCCAAGCCCTGTATTTGATTTG 1082
DB 1026 CTTGGGCTGACCGAAGCTTGAATGTGAGGTTACGACCCCAAGCCCTGTATTTGATTTG 1085
QY 1083 CCGCACCTTCAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAAGATGTTCA 1142
DB 1086 CCGCACCTTCAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAAGATGTTCA 1145
QY 1143 TGGGATTTGTGGGCAATTTCTGAGGCTGAGGAGGCTTTTGAAGAAAGAGAGGA 1202
DB 1146 TGGGATTTGTGTGGCAATTTCTGAGGCTGAGGAGGCTTTTGAAGAAAGATGAGGA 1205
QY 1203 CTATATCTGCGCAAACTGACCAATTTCTGCAAGTGCAGATGAGACTCATTCAGAAAGCG 1262
DB 1206 CTATATCTGCGCAAACTGACCAATTTCTGCAAGTGCAGATGAGACTCATTCAGAAAGCG 1265
QY 1263 AGATCAGCAGAGAGCTAAATGAGACCTGAGAGTGTGATGAGCCGATTTGACAGAT 1322
DB 1266 AGATCAGCAGAGAGCTAAATGAGACCTGAGAGTGTGATGAGCCGATTTGACAGAT 1325
QY 1323 AGGAACAATGAGAGAGAGTCTAGCGAAGACCAAGGATTAAGGATTAAGATTAAGAAAG 1382
DB 1326 AGGAACAATGAGAGAGAGTCTAGCGAAGACCAAGGATTAAGGATTAAGATTAAGAAAG 1385
QY 1383 TGCATATTCAGATGCGCAAGAAACCTCAGATCTTCCAGCTGTGTGATGAGGCGCTG 1442
DB 1386 TGCATATTCAGATGCGCAAGAAACCTCAGATCTTCCAGCTGTGTGATGAGGCGCTG 1445
QY 1443 TGCTTCAAAATGATTTGGGCGCGGCTGTCAAGTGGCGGAGCCGACTGCTGTACTG 1502
DB 1446 TGCTTCAAAATGATTTGGGCGCGGCTGTCAAGTGGCGGAGCCGACTGCTGTACTG 1505
QY 1503 CAGTATGACTGTATCTCAACACAGCGCGCAGCACAATGAAGTTTCTAAGCTCAGTTAA 1562
DB 1506 CAGTATGACTGTATCTCAACACAGCGCGCAGCACAATGAAGTTTCTAAGCTCAGTTAA 1565
QY 1563 AGAA CAGAGCCCAAGCTTAAGAAAGATGAAGTGAAGCCAGAGAGCCCAAGCTTCC 1622
DB 1566 AGAA CAGAGCCCAAGCTTAAGAAAGATGAAGTGAAGCCAGAGAGCCCAAGCTTCC 1625
QY 1623 GAAATGCGGTCTCAGAGGAGTATTAAGTCTTCTGTGCAAGAGACCAAGCTCCAGA 1682
DB 1626 GAAATGCGGTCTCAGAGGAGTATTAAGTCTTCTGTGCAAGAGACCAAGCTCCAGA 1685
QY 1683 AAAAAAGAGACACAGTGAAGAGAGCAGTGTGTCTCTGCGCGGAGTGAAGCACTCG 1742
DB 1686 AAAAAAGAGACACAGTGAAGAGAGCAGTGTGTCTCTGCGCGGAGTGAAGCACTCG 1745
QY 1743 GAAAGAGACGCTTGTGAGAGACGACGCGCTGTGTGGCGAGGATTCATTTCAATTC 1802
DB 1746 GAAAGAGACGCTTGTGAGAGACGACGCGCTGTGTGGCGAGGATTCATTTCAATTC 1805
QY 1803 AGTAAGCCAGAAAGAAAGTGTGCTCTGCGCGGCTCACTGTGTATTAATGATATCA 1862
DB 1806 AGTAAGCCAGAAAGAAAGTGTGCTCTGCGCGGCTCACTGTGTATTAATGATATCA 1865
QY 1863 CCTAAGGTTGGCTCTCTGAGACCCCTCCGTTCTTCTGATATGATATGATATGATATG 1922
DB 1866 CCTAAGGTTGGCTCTCTGAGACCCCTCCGTTCTTCTGATATGATATGATATGATATG 1924
QY 1923 TCCAGAGCTGGAAGTGTGAGCTTTGTGTAAGCTGATCAAGAGACCGGCTGACATCA 1982
DB 1925 TCCAGAGCTGGAAGTGTGAGC-TTGTGTAAGCTGATCAAGAGACCGGCTGACATCA 1983
QY 1983 GCGG 1986
DB 1984 GCGG 1987
```



```

Db      1266 AGATCAGCAGAGAGCTAATGAGACCTGAGATGCTGATGACACGATGTGACAGTAT 1325
QY      1323 AGGAAACATGAGCAGAAAGTCTACCGAAGACCAAGGATTAAGGATTAAGGAAAGC 1382
Db      1326 AGGAAACATGAGCAGAAAGTCTACCGAAGACCAAGGATTAAGGATTAAGGAAAGC 1385
QY      1383 TGCAGAAATCCAAAGTGGCAGAAAGAACTCAAGATCTTCAGAGCTGTGTAGAGGCGCTGG 1442
Db      1386 TGCAGAAATCCAAAGTGGCAGAAAGAACTCAAGATCTTCAGAGCTGTGTAGAGGCGCTGG 1445
QY      1443 TGCCTCAAAAATGTAATGAGCCCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1502
Db      1446 TGCCTCAAAAATGTAATGAGCCCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1505
QY      1503 CAGTAATGACTGTATCTTCAAAACAGCCGCGACGCAATGAAGTTTCTAAGCTCAGGTAA 1562
Db      1506 CAGTAATGACTGTATCTTCAAAACAGCCGCGACGCAATGAAGTTTCTAAGCTCAGGTAA 1565
QY      1563 AGAACAAGCCAAAGCCTAAAGAAAGATGAGAGTGAAGCCAGAGCCAGCTTCC 1622
Db      1566 AGAACAAGCCAAAGCCTAAAGAAAGATGAGAGTGAAGCCAGAGCCAGCTTCC 1625
QY      1623 GAAATGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1682
Db      1626 GAAATGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1685
QY      1683 AAAAAAGAGACCAAGTGAAGAAAGGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1742
Db      1686 AAAAAAGAGACCAAGTGAAGAAAGGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1745
QY      1743 GAAGGAAAGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1802
Db      1746 GAAGGAAAGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1805
QY      1803 AGTAAGCCAGAAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1862
Db      1806 AGTAAGCCAGAAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1865
QY      1863 CTTTGGGGTGGCTCTGTGACCCCTCCCGTTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1922
Db      1866 CTTTGGGGTGGCTCTGTGACCCCTCCCGTTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1924
QY      1923 TCCAGGACTGGGAGTGTGAGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1982
Db      1925 TCCAGGACTGGGAGTGTGAGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1983
QY      1983 GCGG 1986
Db      1984 GCAG 1987

```

RESULT 7
 AAC77933
 ID AAC77933 standard; cDNA; 1764 BP.

XX AAC77933;

DT 08-FEB-2001 (first entry)

DE Human cancer associated gene sequence SEQ ID NO:327.

XX Human; cancer associated gene; cancer antigen; detection; cancer;
 XX diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
 XX antidiabetic; antineoplastic; antithrombotic; antitumor; antiviral;
 XX antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiac;
 XX dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
 XX vasotropic; antipsoriatic; angiogenic; gene therapy; inflammation;
 XX immune disorder; haematopoietic cell disorder; autoimmune disorder;
 XX allergic reaction; graft versus host disease; organ rejection;
 XX haemostatic; thrombolytic; cardiovascular disorder; infection;
 XX neurological disease; drug screening; ss.
 OS Homo sapiens.

```

XX      MO200055350-A1.
PN      21-SEP-2000.
XX      08-MAR-2000; 2000MO-US005882.
PD      12-MAR-1999; 99US-0124270P.
PP      (HUMA-) HUMAN GENOME SCI INC.
XX      Rosen CA, Ruben SM;
XX      MPI: 2000-587533/55.
DR      P-PSDB: AAB43724.
PT      Novel isolated nucleic acids comprising sequences encoding peptides
PT      useful for treating or diagnosing e.g. cancer.
PS      Claim 1; Page 880-881; 2352pp; English.
XX      AAC77607 to AAC78448 encode the human cancer associated proteins given in
XX      AAB43398 to AAB44239. The proteins can have activities based on the
XX      tissues and cells the genes are expressed in. Example of activities
XX      include: cytostatic; proliferative; vulnary; immunomodulator;
XX      antidiabetic; antineoplastic; antithrombotic; antitumor; antiviral;
XX      antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiac;
XX      dermatological; neuroprotective; thrombolytic; coagulant;
XX      neotropic; vasotropic; antipsoriatic and antitumor. The
XX      polynucleotides and polypeptides can be used for preventing, treating or
XX      ameliorating medical conditions and diagnosing pathological conditions.
XX      Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX      the present invention may be used to treat immune disorders by activating
XX      or inhibiting the proliferation, differentiation or mobilisation of
XX      immune cells, to treat disorders of haematopoietic cells, autoimmune
XX      disorders, allergic reactions, graft versus host disease and organ
XX      rejection, modulate haemostatic or thrombolytic activity, modulate
XX      inflammation, cancers, cardiovascular disorders, neurological disease and
XX      bacterial or viral infections. The peptides, nucleotides, antibodies,
XX      agonists and antagonists may be also be used in drug screens. AAC78449 to
XX      AAC78457 and AAB44240 represent sequences used in the exemplification of
XX      the present invention

```

Sequence 1764 BP; 468 A; 424 C; 415 G; 453 T; 0 U; 4 Other;

Query Match 59.9%; Score 1564.4; DB 3; Length 1764;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1610; Conservative 0; Mismatches 2; Indels 5; Gaps 4;

```

QY      996 GGAACATCAAGATGAGAGAGCTGTGAGAC-TTGGGCGGACCGAAGCTGATGTGAGGTT 1054
Db      1 GGAACATCAAGATGAGAGAGCTGTGAGACTTTTGGGCGGACCGAAGCTGATGTGAGGTT 60
QY      1055 ACGACCCCAAGCGCCTGTATTGTCATTTGCCCGCAGCTTCACAAACAGGTTTATGATTT 1114
Db      61 ACGACCCCAAGCGCCTGTATTGTCATTTGCCCGCAGCTTCACAAACAGGTTTATGATTT 120
QY      1115 GCTGTGACCGCTGTGAAAGATGTTTTCATGAGGATTTGTGTGGCATTTCTGAGGCTCGAG 1174
Db      121 GCTGTGACCGCTGTGAAAGATGTTTTCATGAGGATTTGTGTGGCATTTCTGAGGCTCGAG 180
QY      1175 GGAGGCTTTTGGAAAGAGATGGGGAAGACTATATCTGCGCAACTGCAACATTCTGCAAG 1234
Db      181 GGAGGCTTTTGGAAAGAGATGGGGAAGACTATATCTGCGCAACTGCAACATTCTGCAAG 240
QY      1235 TGCAGAGATGAGCTCATTTCAAGAAACGCGCAATCAGAGGAGCTTAATGAGAACCTGAG 1294
Db      241 TGCAGAGATGAGCTCATTTCAAGAAACGCGCAATCAGAGGAGCTTAATGAGAACCTGAG 300
QY      1295 ATGCTGATGCGCACCGATTTGTAAGATTAAGAAACATGAGAGAGTGTAGCGAAGACC 1354
Db      301 ATGCTGATGCGCACCGATTTGTAAGATTAAGAAACATGAGAGAGTGTAGCGAAGACC 360

```


QY 1355 AAGGATTAAGGTTGATTAAGTGAAGAGCTGCATAATCCAAATGCGCAAGAGAACTCAAGA 1414
DB 361 AAGGATTAAGGTTGATTAAGTGAAGAGCTGCATAATCCAAATGCGCAAGAGAACTCAAGA 420
QY 1415 TCTTCCAGCCCTGTGATTAAGGCGCTGGTGCCTCAAAAATGATTTGGCCCCGGGTCTGTG 1474
DB 421 TCTTCCAGCCCTGTGATTAAGGCGCTGGTGCCTCAAAAATGATTTGGCCCCGGGTCTGTG 480
QY 1475 ACGTGGCGCAGCCGACTCGGTTGACTGAGTAAATGACTGATCTCTCAAAACAGCCGAG 1534
DB 481 ACGTGGCGCA-CCGACTCTGGTGTACTGAGTAAATGACTGATCTCTCAAAACAGCCGAG 539
QY 1535 CGACAATGAAGTTTCTAAGCTCAAGTAAAGACAGAACCCAAAGCTTAAAGAAAGATGA 1594
DB 540 CGACAATGAAGTTTCTAAGCTCAAGTAAAGACAGAACCCAAAGCTTAAAGAAAGATGA 599
QY 1595 AGATGAAGCCAGAGAGCCCACTCTCCGAAATGGCGGTCTCAAGCCAGGTATTAATCT 1654
DB 600 AGATGAAGCCAGAGAGAGCCCACTCTCCGAAATGGCGGTCTCAAGCCAGGTATTAATCT 659
QY 1655 CTTCTGTGCAAGAGAGCCAGCTCCAGAAAGAAAGACCAAGTGAAGAGCAGTGG 1714
DB 660 CTTCTGTGCAAGAGAGCCAGCTCCAGAAAGAAAGACCAAGTGAAGAGCAGTGG 719
QY 1715 TGGTCCCTGCGCGAGTGAAGCACTCGGGAAGAAAGCAGCTTGTGAAGCAGCAGCTG 1774
DB 720 TGGTCCCTGCGCGAGTGAAGCACTCGGGAAGAAAGCAGCTTGTGAAGCAGCAGCTG 779
QY 1775 CGTGGCGAGAGATGCAATTAATGCAATGCAATGAACCCAGAAAGATGCTGCTCCCTG 1834
DB 780 CGTGGCGAGAGATGCAATTAATGCAATGCAATGAACCCAGAAAGATGCTGCTCCCTG 839
QY 1835 CGTCACTGTGTATTAATGATATGATCACTAGGGGTGGCTCTCTGAGCCCTCCCTG 1894
DB 840 CGTCACTGTGTATTAATGATATGATCACTAGGGGTGGCTCTCTGAGCCCTCCCTG 899
QY 1895 CTTTCTGTGATNAGCCATCCCTGAGGCTGTGTCAGAGCTGGAGTTGACCTTTGTGTAAG 1954
DB 900 CTTTCTGTGATNAGCCATCCCTGAGGCTGTGTCAGAGCTGGAGTTGACCTTTGTGTAAG 959
QY 1955 CTGATCAAGACACCGGCTGACCACTCAAGCCGGAAGCAGAGCCCAATGTCAGAGATGCC 2014
DB 960 CTGATCAAGACACCGGCTGACCACTCAAGCCGGAAGCAGAGCCCAATGTCAGAGATGCC 1019
QY 2015 CTGCTGCTGTGATCCCTGATCCCTGATGCTGACAGACTTCTCTGCTCAAGCTG 2074
DB 1020 CTGCTGCTGTGATCCCTGATCCCTGATGCTGACAGACTTCTCTGCTCAAGCTG 1079
QY 2075 TAAACCTGACTGTGTAAGCGTTCACTTAATGATTAATCTTTAATCTGTTTCACTCT 2134
DB 1080 TAAACCTGACTGTGTAAGCGTTCACTTAATGATTAATCTTTAATCTGTTTCACTCT 1139
QY 2135 CAGGCTCTGTGTAAGTATTTGATTTCTTTCAATCCAGCTGTGATGATGCAAGCTGCC 2194
DB 1140 CAGGCTCTGTGTAAGTATTTGATTTCTTTCAATCCAGCTGTGATGATGCAAGCTGCC 1199
QY 2195 CGGCAAGCCACATCCACCCTGTCTGACATGAGTTGTTCTGACAAAGCGCTGTATAG 2254
DB 1200 CGGCAAGCCACATCCACCCTGTGTCTGACATGAGTTGTTCTGACAAAGCGCTGTATAG 1259
QY 2255 CTTGAGTTTTCACATTTGTCCACGCGCCAGACATGAAGATCACTTTTATATGTT 2314
DB 1260 CTTGAGTTTTCACATTTGTCCACGCGCCAGACATGAAGATCACTTTTATATGTT 1319
QY 2315 GTGGGAATCTTTGCAAGTTAGTTGATTTCTTTCAATCCAGCTGTGATGATTTATGACT 2374
DB 1320 GTGGGAATCTTTGCAAGTTAGTTGATTTCTTTCAATCCAGCTGTGATGATTTATGACT 1379
QY 2375 GGGGAGATAGGGGATTTTTTTTTTCCATGTCGATTCACACCTCAACACCCACATG 2434
DB 1380 GGGGAGATAGGGGA--TTTTTTTTTCCATGTCGATTCACACCTCAACACCCACATG 1437
QY 2435 AACAATTCGAATTCGAAGG-CACACACTCTGCTTCATAGAGCCCAAGTAAAGT 2493

DB 1438 AACACATTCGAATTCGAAGG-CACACACTCTGCTTCATAGAGCCCAAGTAAAGT 1497
QY 2494 TCACACTAGAACACTGTCTGACCGCAGAGCGCTGCTTGACTGTGATTTCAATG 2553
DB 1498 TCACACTAGAACACTGTCTGACCGCAGAGCGCTGCTTGACTGTGATTTCAATG 1557
QY 2554 TGACTGGCTTTCTTGCCCTCGTCTCTTGAAATTTTAACTCTTAAGATATATCCG 2610
DB 1558 TGACTGGCTTTCTTGCCCTCGTCTCTTGAAATTTTAACTCTTAAGATATATCCG 1614
RESULT 8
AAK53352
ID AAK53352 standard; cDNA; 7290 BP.
XX
AC AAK53352;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2881.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
OS Homo sapiens.
XX
PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
XX
PR 27-APR-2000; 2000US-00560875.
XX
PR 20-JUN-2000; 2000US-00598075.
XX
PR 19-JUL-2000; 2000US-00620325.
XX
PR 01-SEP-2000; 2000US-00654936.
XX
PR 15-SEP-2000; 2000US-00663561.
XX
PR 20-OCT-2000; 2000US-00693325.
XX
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HVS-E) HVS-EQ INC.
XX
PI Tang YF, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR MPI; 2001-476283/51.
XX
DR P-PSDB; AAM80219.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 5053-5055; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Notes: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 7290 BP; 1762 A; 1953 C; 1975 G; 1600 T; 0 U; 0 Other;

Query March 49.1%; Score 1280.6; DB 4; Length 7290;
 Best Local Similarity 92.6%; Pred. No. 0;
 Matches 1401; Conservative 0; Mismatches 4; Indels 108; Gaps 1;

455 GTGGAGGAGCAGCCAGGCGACTGAGCGCTGAGCAGCTTCTGACCAATTCGCGCGCGCC 514
 64 GTGGAGGAGCAGCCAGGCGACTGAGCGCTGAGCAGCTTCTGACCAATTCGCGCGCGCC 123
 515 GCGGAGGAGGAGCAGCTTCTGAGGAGATTCGTGTGAGCCGACGCTCTGCGCCG 574
 124 GCGGAGGAGGAGCAGCTTCTGAGGAGATTCGTGTGAGCCGACGCTCTGCGCCG 183
 575 CCAAGAGCGCGGAGAGAGCCCTCCAGGCGCAGGTGAGGAGGCTTGTGAGCAGAGAGC 634
 184 CCAAGAGCGCGGAGAGAGCCCTCCAGGCGCAGGTGAGGAGGCTTGTGAGCAGAGAGC 243
 635 GCGGAGGAGGAGCAGCTTCTGAGGAGATTCGTGTGAGCCGACGCTCTGCGCCG 694
 244 GCGGAGGAGGAGCAGCTTCTGAGGAGATTCGTGTGAGCCGACGCTCTGCGCCG 303
 695 GAGGAGGAGGAGCAGCTTCTGAGGAGATTCGTGTGAGCCGACGCTCTGCGCCG 754
 304 GAGGAGGAGGAGCAGCTTCTGAGGAGATTCGTGTGAGCCGACGCTCTGCGCCG 363
 755 TTCAAGATTCGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 814
 364 TTCAAGATTCGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 423
 815 TTCAAGATTCGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 874
 424 TTCAAGATTCGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 483
 875 AGGCGAGTGAACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 934
 484 AGGCGAGTGAACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 543
 935 GGGTTGTGTCCAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 994
 544 GGGTTGTGTCCAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 603
 995 AGGCGAGTGAACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1054
 604 AGGCGAGTGAACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 663
 1055 AGGCGAGTGAACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1114
 664 AGGCGAGTGAACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 723
 1115 GCTGTGAGCCGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1174
 724 GCTGTGAGCCGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 783
 1175 GAGGAGCTTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1234
 784 GAGGAGCTTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 843
 1235 TGCAAGATGAGACTCATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1294
 844 TGCAAGATGAGACTCATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 903
 1295 ATGCTGATGAGCAGCCATTTGATCAATATATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1354
 904 ATGCTGATGAGCAGCCATTTGATCAATATATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 963
 1355 AAGGAGTAAAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1414
 964 AAGGAGTAAAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1023
 1415 TCTTCAGGCTT----- 1425
 1024 TCTTCAGGCTTCTGCGCGCTGTCCAGCAGCTGCTGTCTGCGAGGATTTGG 1083

1426 -----G 1426
 1084 AAATGTGTGTCTAGAGGAGCTCTGCTTCACTCTTGTGACGATTAAGCTGAGG 1143
 1427 TGAATGAGGCGCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1486
 1144 TGATGAGGCGCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1203
 1487 CCGACTGAGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1546
 1204 CCGACTGAGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1263
 1547 TTCTAAGCTCAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1606
 1264 TTCTAAGCTCAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1323
 1607 AGAAGCCAGTCTTCCGAAATGCGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1666
 1324 AGAAGCCAGTCTTCCGAAATGCGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1383
 1667 AGAAGCCAGTCTTCCGAAATGCGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1726
 1384 AGAAGCCAGTCTTCCGAAATGCGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1443
 1444 GAGTGAAGGAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1503
 1727 GAGTGAAGGAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1786
 1787 ATCAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1846
 1504 ATCAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1563
 1847 ATTAATGATGTA 1859
 1564 ATTAATCAAGAA 1576

RESULT 9
 AAK52368
 ID AAK52368 standard; cDNA; 7838 BP.
 XX
 AC AAK52368;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 913.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HISB-) HISBQ INC.
 XX
 PI Tang YT, Liu C, Dormanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 Ma Y, Zhao Qa, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
DR P-PSDB: AAM79235.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
PS Claim 1, Page 3007-3013; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 7838 BP; 1872 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;
Query Match 49.1%; Score 1280.6; DB 4; Length 7838;
Best Local Similarity 92.6%; Pred. No. 0;
Matches 1401; Conservative 0; Mismatches 4; Indels 108; Gaps 1;
QY 455 GTGGAGGACAGCCCAAGCGCACTGAGCGCGTGGAGCAATTCTGACCAATTGCGCGCGCC 514
DB 599 GTGGAGGACAGCCCAAGCGCACTGAGCGCGTGGAGCAATTCTGACCAATTGCGCGCGCC 658
QY 515 GCGGCGAGAGAGCAATGCTCTGCTCTGCTGAGAGATTTCTGCTGAGCCCAAGTCTCTGCGCCG 574
DB 659 GCGGCGAGAGAGCAATGCTCTGCTCTGCTGAGAGATTTCTGCTGAGCCCAAGTCTCTGCGCCG 718
QY 575 CCACAGAGCGCCGAGCAGCGCTCCGAGGGGCGGTGGAAAGCGCTTCTGAGACAGAAAGCG 634
DB 719 CCACAGAGCGCCGAGCAGCGCTCCGAGGGGCGGTGGAAAGCGCTTCTGAGACAGAAAGCG 778
QY 635 GCCCCAGTCTGCTTCAACAGCTGTGAAGAAAGCAACAGCTCTTCTGAAAAGTGAAG 694
DB 779 GCCCCAGTCTGCTTCAACAGCTGTGAAGAAAGCAACAGCTCTTCTGAAAAGTGAAG 838
QY 695 GAGGGGATGACAGATGACATCTCCGATGAGCAAGGAGGCTGACCTTGAAGAGC 754
DB 839 GAGGGGATGACAGATGACATCTCCGATGAGCAAGGAGGCTGACCTTGAAGAGC 898
QY 755 TTCAAGATCGCTTGGGAGGAGCGGAAAGCAAGAGCCCACTGAGAGGCCCTGAAAAGGGA 814
DB 899 TTCAAGATCGCTTGGGAGGAGCGGAAAGCAAGAGCCCACTGAGAGGCCCTGAAAAGGGA 958
QY 815 TCCAGAGTCCGCTGCGAAGAAAGCGCCGAGAGAGGCTCCGCGAGACTGTGGGCTCCG 874
DB 959 TCCAGAGTCCGCTGCGAAGAAAGCGCCGAGAGAGGCTCCGCGAGACTGTGGGCTCCG 1018
QY 875 AGGCCAGTGAACCTGTGAGAGGCGCTCTGCCAGTGAAGCAAGGCCCGAAGACGATCAG 934
DB 1019 AGGCCAGTGAACCTGTGAGAGGCGCTCTGCCAGTGAAGCAAGGCCCGAAGACGATCAG 1078
QY 935 GGGTTGTCTCCAGGCTGGGAAAGATGACAGAGAGATTAAGTTGAGGGAAGGGGCTC 994
DB 1079 GGGTTGTCTCCAGGCTGGGAAAGATGACAGAGAGATTAAGTTGAGGGAAGGGGCTC 1138
QY 995 AGGACATCAAGAGATGAGAGCTTGAAGACTTGGGCGGACCGAAGCTGATGTGAGGTT 1054
DB 1139 AGGACATCAAGAGATGAGAGCTTGAAGACTTGGGCGGACCGAAGCTGATGTGAGGTT 1198
QY 1055 AGGACCCCAAGCGCTGTATTCATTTGCGCCAGCTTCAAAACAAGGTTTATGATTT 1114
DB 1199 AGGACCCCAAGCGCTGTATTCATTTGCGCCAGCTTCAAAACAAGGTTTATGATTT 1258

QY 1115 GCTGTACCGCTGTGAGAAATGTTTCATGCGGATGTGTGGGCAATTTCTGAGGCTCGAG 1174
DB 1259 GCTGTACCGCTGTGAGAAATGTTTCATGCGGATGTGTGGGCAATTTCTGAGGCTCGAG 1318
QY 1175 GAGGCTTTTGGAAAGAAATGGGAAAGCTATATCTGCCAAATCGACCATTTCTGCAAG 1234
DB 1319 GAGGCTTTTGGAAAGAAATGGGAAAGCTATATCTGCCAAATCGACCATTTCTGCAAG 1378
QY 1235 TCCAGATGAGACTCATTCAGAAAGCGGAGATCAGAGGAAGCTAAATGAGACCTGAG 1294
DB 1379 TCCAGATGAGACTCATTCAGAAAGCGGAGATCAGAGGAAGCTAAATGAGACCTGAG 1438
QY 1295 ATGCTATGCGACCGATTTGTACAGTATAGAAACATAGAGCAAGTCTAGCGAAGCC 1354
DB 1439 ATGCTATGCGACCGATTTGTACAGTATAGAAACATAGAGCAAGTCTAGCGAAGCC 1498
QY 1355 AAGGATTAAGGTTGAATTTGAGAAAGCTGCAATCCAGTGGCAAGAAAGAACTCAAG 1414
DB 1499 AAGGATTAAGGTTGAATTTGAGAAAGCTGCAATCCAGTGGCAAGAAAGAAAGCTCAAG 1558
QY 1415 TCTTCAGCT----- 1425
DB 1559 TCTTCAGCTGTGTCGGGAGCTGTCCGACCCAGCTGCTCTTGGCAGATTTGG 1618
QY 1426 -----G 1426
DB 1619 AATTGTGTGTCTAGAAAGCATCTGCGCTTCAACACTTGTGACATGATAGTCGAG 1678
QY 1427 TGAATGAGCGCTGTGCTGCTCAAAATGATTTGGCCCGGTGCTGACGTGGCCAGC 1486
DB 1679 TGAATGAGCGCTGTGCTGCTCAAAATGATTTGGCCCGGTGCTGACGTGGCCAGC 1738
QY 1487 CCGACTCGGTGTAATGAGTATGATCTGATCTCAAAACAGCGCGAGGAGCAATGAGT 1546
DB 1739 CCGACTCGGTGTAATGAGTATGATCTGATCTCAAAACAGCGCGAGGAGCAATGAGT 1798
QY 1547 TTCTAAGCTCAGGTAAAGAAAGCAAGAGCCAAAGCTTAAAGAAAGATGAAAGTAAAGCCAG 1606
DB 1799 TTCTAAGCTCAGGTAAAGAAAGCAAGAGCCAAAGCTTAAAGAAAGATGAAAGTAAAGCCAG 1858
QY 1607 AGAGCCCAAGTCTTCCGAAATCGGTGCTCAGGCAAGTATTAATCTCTGTGCA 1666
DB 1859 AGAGCCCAAGTCTTCCGAAATCGGTGCTCAGGCAAGTATTAATCTCTGTGTGCA 1918
QY 1667 AAGAGCCAGCTTCCGAAATCGGTGCTCAGGCAAGTATTAATCTCTGTGTGCA 1726
DB 1919 AAGAGCCAGCTTCCGAAATCGGTGCTCAGGCAAGTATTAATCTCTGTGTGCA 1978
QY 1727 GAGTGAAGCACTCGGAAAGAAAGCAAGTGTGAGAGCAAGCGCGTGTGGGCGAGCG 1786
DB 1979 GAGTGAAGCACTCGGAAAGAAAGCAAGTGTGAGAGCAAGCGCGTGTGGGCGAGCG 2038
QY 1787 ATCACAATTACATGCAAGTAAAGCAAGAAAGCAAGTGTGCTCTGCGCGTCACTGTTGT 1846
DB 2039 ATCACAATTACATGCAAGTAAAGCAAGAAAGCAAGTGTGCTCTGCGCGTCACTGTTGT 2098
QY 1847 ATTAATGTATGTA 1859
DB 2099 ATTAATGCCAGAA 2111
RESULT 10
AAS86576 standard; cDNA; 2332 BP.
XX AAS86576;
AC 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #22380.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS
PN MO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX P-PSDB; ABG22389.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 22380; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostic assays as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS6197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;
Query Match 45.6%; Score 1190.8; DB 5; Length 2332;
Best Local Similarity 93.2%; Pred. No. 4.3e-310;
Matches 1362; Conservative 0; Mismatches 37; Indels 79; Gaps 7;
QY 244 GCTTCAGTGTGAGACAGGTATGACGACAAAGCCGAGCATGAGAGGCACT 303
DB 370 GCTTACACTTACCGCTAGGTATGACGACAAAGCCGAGCATGAGAGGCACT 429
QY 304 AAGGCGATCAAAACCCAGGAGAAAGTTGAGAAATAGGGGTTTTCAGAGGCACT 363
DB 430 AAGGCGATCAAAACCCAGGAGAAAGTTGAGAAATAGGGGTTTTCAGAGGCACT 489
QY 364 ATCCCAAGCAGAGGCGCAGAGGAGCGGAGGCTGACCACTGAGCCGCCACCCCA 423
DB 490 ATCCCAAGCAGAGGCGCAGAGGAGCGGAGGCTGACCACTGAGCCGCCACCCCA 549
QY 424 CAGAGAGAGCTGGGCTGTCTCTTGGGCGGCGAGTGGAGGCGAGCCAGCCGACCTGAGGCG 483
DB 550 CAGAGAGAGCTGGGCTGTCTCTTGGGCGGCGAGTGGAGGCGAGCCAGCCGACCTGAGGCG 609
QY 484 GTGAGAGAGTTCTGACCAATTGGCGGCGGCGGAGGAGAGAGAGTCTGTCTCCCTG 543

DB 610 GTGAGAGAGTTCTGACCAATTGGCGGCGGCGGAGGAGAGAGAGTCTGTCTCCCTG 669
QY 544 GAGGATCTGGTGAAGCCCAAGTCTTGCCTCCGCGCACAAGCGCGAGAGCCCTCCGAGGCG 603
DB 670 GAGGATCTGGTGAAGCCCAAGTCTTGCCTCCGCGCACAAGCGCGAGAGCCCTCCGAGGCG 729
QY 604 AGCGTGAAGAGCGCTTCTGAGACCAAGAGCGGCCCCAGCTGCTTCCACAGCTGTGAAG 663
DB 730 AGCGTGAAGAGCGCTTCTGAGACCAAGAGCGGCCCCAGCTGCTTCCACAGCTGTGAAG 789
QY 664 GAACGACCAAGCTCTTCTGAAAAAGGTGAAGAGAGGATGACCAAGATGACCTTCGAT 723
DB 790 GAACGACCAAGCTCTTCTGAAAAAGGTGAAGAGGATGACCAAGATGACCTTCGAT 849
QY 724 AGTGAACGCGATGGCTGACCTTGAAGAGTTCAGAAATGCGCTTCCGAGAGCGGGA 783
DB 850 AGTGAACGCGATGGCTGACCTTGAAGAGTTCAGAAATGCGCTTCCGAGAGCGGGA 909
QY 784 CAGAGGCCCACTGAGAGGCGCCCTGAAGAGGATCCAGAGTCCGCTCCGAGAGAGCGCGG 843
DB 910 CAGAGGCCCACTGAGAGGCGCCCTGAAGAGGATCCAGAGTCCGCTCCGAGAGAGCGCGG 969
QY 844 GAGGAGGCTCCGCGCGAGACTGTGAGGCTCCGAGGCGCACTGTCGAGAGGCTCTTG 903
DB 970 GAGGAGGCTCCGCGCGAGACTGTGAGGCTCCGAGGCGCACTGTCGAGAGGCTCTTG 1029
QY 904 CCGATGAAGCAGAGGCGCCGAGAGAGATCAAGGCGGTGTCTCCAGGCTGGGAAAGATAC 963
DB 1030 CCGATGAAGCAGAGGCGCCGAGAGAGATCAAGGCGGTGTCTCCAGGCTGGGAAAGATAC 1089
QY 964 AGAGAGGTAAAGTTGAGAGGAGGAGGAGGCTCAGAGCATCAAGATGAGAGGCTGAGAC 1023
DB 1090 AGAGAGGTAAAGTTGAGAGGAGGAGGAGGCTCAGAGCATCAAGATGAGAGGCTGAGAC 1149
QY 1024 -TTGGGCGCGAGCGAGCTGAATGAGAGGTTACGACCCCAAGCTCTGTATTCATTG 1082
DB 1150 TTGGGCGCGAGCGAGCTGAATGAGAGGTTACGACCCCAAGCTCTGTATTCATTG 1209
QY 1083 CCGCGAGCTTCAACAACAAGGTTTATGATTTGCTGAGCCGCTGTGAAGATGGTTCA 1142
DB 1210 CCGCGAGCTTCAACAACAAGGTTTATGATTTGCTGAGCCGCTGTGAAGATGGTTCA 1269
QY 1143 TGGGAGTGTGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGAAATGGGAGA 1202
DB 1270 TGGGAGTGTGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGAAATGGGAGA 1329
QY 1203 CTATATCTGCCCAACTGCAACATTTCTGCAAGTGCAGAGATGACATTCAGAAACGCG 1262
DB 1330 CTATATCTGCCCAACTGCAACATTTCTGCAAGTGCAGAGATGACATTCAGAAACGCG 1389
QY 1263 AGATCAGCAGAGAGCTTAATGAGAG---CTTGAAGATGC-TGATGGACCGA-TTGTACA 1317
DB 1390 AGATCAGCAGAGAGCTTAATGAGAGAGCTTGGAGATGCTTGAATGAGCCGATTTTACA 1449
QY 1318 AGTATGAGAAACAT--AGAGCAGAGTCT-----AGCGAAGCAAGGATTAAGGATAGA 1371
DB 1450 AGTATGAGAAACATTAAGAGAGAGAGTCTTAAGCGAAGACCCAGAGGTTTAAAGGTAGA 1509
QY 1372 ATTGAAGAAAGCTGCAAAATCCAAAGTGGCAGAGAAACTCAAGATTTCTCAAGC----- 1424
DB 1510 ATTGAAGAAAGCTGCAAAATCCAAAGTGGCAGAGAAACTCAAGATTTCTCAAGCCTGATTT 1569
QY 1425 ----- 1424
DB 1570 GGAATATGCTGTGTCTGAAGACATCTGCTTCAACCTTTCAGACTGACATTAAGTGCAC 1629
QY 1425 TGTGATGAGAGCGCTGCTGCTGCTCAAAAATGATTTGGCCCGGAGTGTCTCACTGAGCCCA 1484
DB 1630 GGTGATGAGAGCGCTGCTGCTGCTCAAAAATGATTTGGCCCGGAGTGTCTCACTGAGCCCA 1689
QY 1485 GCCGAGCTCGGTGTACTGCAATTAATGACTGTATCTTCAACAGCGCGCAGGAGCAATGAA 1544
DB 1690 GCCGAGCTCGGTGTACTGCAATTAATGACTGTATCTTCAACAGCGCGCAGGAGCAATGAA 1749

Qy	1545	GTCTTAAAGCTCAGGTAAAGAAAGAACCCAAAGCCTTAAAGAAAAGATGAAGTGAAGCC	1604
Db	1750	GTCTTAAAGCTCAGGTAAAGAAAGAACCCAAAGCCTTAAAGAAAAGATGAAGTGAAGCC	1809
Qy	1605	AGAGAACCCAGTCTTCGAAATGCGGTCTCAGGACG	1642
Db	1810	AGAGAACCCAGTCTTCGAAATGCGGTCTCAGGCTG	1847
RESULT 11			
ID	AA290579		
XX	AA290579	standard: cDNA; 2867 BP.	
XX	AA290579;		
XX	19-JUN-2000	(first entry)	
DE	Murine death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.		
XX	XX	Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; murine;	
KM	XX	autoimmune disease; cytostatic; immunosuppressive; antidiabetic;	
KM	XX	antirheumatic; antiinflammatory; antiproliferative; ss.	
OS	XX		
XX	XX	Mus sp.	
XX	XX		
XX	XX	Location/Qualifiers	
FT	Key	232..2076	
FT	CDS	/*tag= a	
FT	FT	/product= "DIO-1"	
PN	MO200015767-A1.		
XX	XX		
PD	23-MAR-2000.		
XX	XX		
PP	10-SEP-1999;	99WO-GB003019.	
XX	XX		
PR	10-SEP-1998;	98SR-00003069.	
XX	XX		
PR	17-SEP-1998;	98US-0100873P.	
XX	XX		
XX	(CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.		
PA	(BANN/) BANNERMAN D G.		
PI	Alonso CM, Domingo DG, Grandlen A, Leonardo B, Martinez P;		
XX	XX		
DR	XX	WPI: 2000-271426/23.	
DR	XX	P-PSDB; AAY67580.	
XX	XX		
PT	New DNA encoding human and murine death inducer-obliterator 1		
PT	polypeptides, useful in the treatment of cancer, autoimmune diseases,		
PT	diabetes, rheumatoid arthritis, benign tumors, malignant tumors and		
PT	hyperproliferative skin disorders.		
XX	XX		
XX	XX	Claim 3; Fig 1B; 27pp; English.	
XX	XX		
CC	The invention provides nucleic acids encoding the human and murine death		
CC	inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be		
CC	expressed by standard recombinant methodology. The DIO-1 polypeptides,		
CC	agonists and antagonists are used as a medicament for treating diseases		
CC	characterized by an alteration in cell death or by hyperproliferation,		
CC	e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign		
CC	tumours, malignant tumours or hyperproliferative skin disorders. They are		
CC	also useful in the treatment of metabolic, proliferative or inflammatory		
CC	conditions. The present sequence represents a cDNA encoding the murine		
CC	DIO-1 polypeptide		
XX	XX		
SQ	Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;		
XX	XX		
Query Match	39.2%;	Score 1023.6; DB 3; Length 2867;	
Best Local Similarity	71.9%;	Pred. No. 5.9e-265;	
Matches 1494; Conservative	0; Mismatches 534;	Indels 50; Gaps 10;	
21	TCGCGGCGGCTTCGGGGAATGGCTGCGAGACCTTGAAGCCTGCGGACCTTACTTCACGG 80		

Db 1 TCCGTGTAGCTCTGGAAATGGCTGGCGGATCCCGGGGCGGGAGCTTGTTAAGAGG 60
 Oy 81 GAAGAGCCTCTAGATAATCTGAAGTTGTGAAAAATACAGAGCCTGTTACTGTGAACAGTG 140
 Db 61 CAGTCCCACTGTCCCTTGTGGTTGTGAAAGCTCCGGAATCTTCTCATGTGAACGTGTG 120
 Oy 141 GCTGACAAAGTGTGTTGTGAGGCTGGCTGTCTGCTTGGACCCAGAGGTTTGTGTCTGCC 200
 Db 121 ACTGACAAACAGTGGGGTAG--GCTTGGCCGTCTGCTGAAC-----TGGCCC 166
 Oy 201 AGGGTTTTGGTGTATTTAGATTT-CAGGGAAAGTGTCCAGGCTTTCAGTGTGGAG 259
 Db 167 CAGGTCTATATATTTATATGTAGAGATTTCACCCAAAGGTTTCCAGGCTTTCAGTTTGGGA 226
 Oy 260 CAGGTATGGAAGCAAAAGCGAGCCCGAGCAATGAGAGGCACTTAAGGCCATCAAAACCA 319
 Db 227 CAGGTATGATGATTAAGAGGCACTGAGCAATAGAGAGGACCCCAAGGCTATCAAAACCA 286
 Oy 320 CCAGCAAAAGTTCCAGAAAACATGGGGTTTGGAGGACCACTATTCGCAAGCGAGAG 379
 Db 287 CCAGTAAAGAGTTCCAGAAAACCTGGGGTTTGGAGAACCAAGATTGCCAAACGTGAGG 346
 Oy 380 GCGAGGGGACCGAGAGGCTGACCCCACTGGAAGCCGACCCCAAGCAGCAGCTGGGCC 439
 Db 347 GTGAGAGAGACAGAGAGGCGAGACCCCACTAGACAGCAAC-----ACACAGCATTAAC 400
 Oy 440 TGTCCCTGCGGCGCAGTGGAGGCGAGCCCAAGCGCACTGAGCGCTGAGCAGTTCCTGA 499
 Db 401 TCTCCCTGCGCGAGTGAAGCGCAACAAAGTACTGAGAGAGGTGAGAGTTCTTA 460
 Oy 500 CCAATTGCGCGCGCCGCGAGAGAGACATGCCCTGTCTCCCTGAGAGATTCTGTGAGC 559
 Db 461 CCAAGGTTCCGCGCGCAGAGGAAAAAGATGTGCCGCGTGTCCCTGAGAGATTCCAGTAGC 520
 Oy 560 CCAGTCTCTGCCCCCGACAGAGCGCGAGCAGCTCCGAGGGGCAAGGTGGAAGAGGCTT 619
 Db 521 CCACATCTTCCACAGTACATGATGTGAGACAGCTTCCGAGGGAGGCTTGAAGACGTT 580
 Oy 620 CTGAGACCAAGAAAGGCGCCCAAGCTGCTTCCACAGCTGTGAAGAAAGCAAGCAGCTCTT 679
 Db 581 CTGAGATCAGAAAGTGGCCCTGTATCTGATCTCTTA---GGAAAGAACATCTGCGCTCTT 637
 Oy 680 CTGAAAGGTGAAGAGAGGGAGATGACCAAGATGACATCTCCATATGACAGCGATGGC 739
 Db 638 CTGAAAGGCAAAAGAGGTGAAGAGAAAGAACACTCTGACAGTGAAGATGGCC 697
 Oy 740 TGACTCTTGAAGAGCTTCAAGATTCGCTTGGCAGGAAGCGGGAACAGGAGCCCACTGAGA 799
 Db 698 TTAGGTGAAGGAATCTTCAAGACCGCTTGGAGAAAGCCAGAGCAAGAACCTGTGAGGA 757
 Oy 800 GGCCCTGAAAGGGAATCCAGAGTCCGCTGGGAGAAAGCGCCGGGAGAGGGTCCCGCG 859
 Db 758 GGTCCCTGAAGAGGCAATCAAGATTCGCTTGAAGAAAGAGCCCAAGAGAGAGATTCTGCC 817
 Oy 860 AGACTGTGTGGCTCCGAGGCAAGTGAACATGTGAGAGGGCGTCTTGGCCAGTAAAGAGAGC 919
 Db 818 AAATCTGGGAGTGTCCAAATATGGCAGTGCAGAGCAGGACAGACTCTCTGTATACAGGAGC 877
 Oy 920 CCGAAGAACATCAGGGGGTTGTGTCCGAGCTGGGAAGATGACAGAGAGATTAATTTGG 979
 Db 878 CTGAGGCTAGTCAAGGAGCAGGTGTCCCAAGTCAAGAGCAGATGACATTAAGAAATCAATTTG 937
 Oy 980 AGGGAAGGCGGCTCAGAGACATCAAAAGATGAGAGGCTTGGAGACTTGGGCGCACCGAAGC 1039
 Db 938 AAGGAAGGCGCACTCAAGGAATATCAGAGGAAGAAACCCCAAGGAGCGGGCAACCAAGC 997
 Oy 1040 CTGAATGTGAAGGTTACGACCCCAAGCCGCTGTATATTCATTTGCCGCAAGCTTCAACA 1099
 Db 998 CTGAATGTGAAGGTTTACGACCCCAAGTCCCTGTATCTGCACTGCGGCAAGCTTCAACA 1057
 Oy 1100 ACAGGTTTATGATTTGCTGTGAACGCTGTGAAGAAATGTTTCAATGCGGATTTGTGTGGCA 1159

Db 1058 ACAGTTTATGATCTGCTGATGATGCTGTGAGAGATGTTCCATGCTGACTGTGGGTA 1117
Qy 1160 TTTCTGAGGCTCCAGGAGGCTTTTGGAAAAGATGGGGAGATATATCTGCCAAACT 1219
Db 1118 TTTCTGAGGCTCCAGGAGGCTCTCTGAAAGAAAGGAGGAGCTACATCTGCCAAAT 1177
Qy 1220 GCACATTTCTGCAAGTACAGATAGATGATCTCATTCAGAAAACGAGCATCAGCAGAGCTA 1279
Db 1178 GCACATTTTTCAGATGTCAGATGATGATGATGATGATGATGATGATGATGATGATG 1237
Qy 1280 AATGAGACCTGAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1339
Db 1238 GGTGACATCTGTGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1297
Qy 1340 AGTCTAGCAGAAAGCAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1399
Db 1298 AGTCCGAGAAAGCAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1357
Qy 1400 AGAAGAACTCAAGATCTTCCAGCCTGTGATGATGATGATGATGATGATGATGATGATG 1459
Db 1358 AGAAGAACTCAAGATCTTCCAGCCTGTGATGATGATGATGATGATGATGATGATGATG 1417
Qy 1460 GCCCCGGGCTCTGCTACGCTGCGAGCCGACCTGCTGATGATGATGATGATGATGATGATG 1519
Db 1418 GCCCTGGGTGTTCAGTGTAGCAAGCCTGATCTGTGTATGATGATGATGATGATGATGATG 1477
Qy 1520 TCAAAACGCGCGCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1579
Db 1478 TCAAAACGCGCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1537
Qy 1580 CTAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1639
Db 1538 CTAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1597
Qy 1640 CAGTATTAATAATCTCTCTGTGCAAGAGACCAAGCTCCAGAAAAAAGAGACCAAG 1699
Db 1598 TGGGAGTTAAATCTCTCTGTGCAAGAGACCAAGCTCCAGAAAAAAGAGACCAAG 1657
Qy 1700 TGAAGAAAGCAGTGTGTGCTCTGCGAGATGAAAGCACTGGGAGAGAGAGCACTTGTG 1759
Db 1658 TGAAGAAAGCAGTGTGTGCTCTGCGAGATGAAAGCACTGGGAGAGAGAGCACTTGTG 1714
Qy 1760 AGAGCAGACGCGCTGCGAG 1819
Db 1715 AGAGCAGACGCGCTGCGAG 1774
Qy 1820 CTG-----CTGCTCCCTGCGCGCTCACTGTGTATTAATGTATGTATCACTAGGGG 1870
Db 1775 CAGAGAAAGCCCACTGCACTCTGCGCAAGCCATTTGAGTAATGTATCACTAGGGG 1834
Qy 1871 TTGGCTCTCTGAGACCTCTCCGTTCTTCTGATAGCCATCCCTGGGCTGTCTAGAGAC 1930
Db 1835 CTGGCTCTCTGAGACCTCTCCGTTCTTCTGATAGCCATCCCTGGGCTGTCTAGAGAC 1889
Qy 1931 TTGGAGTTGACGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1990
Db 1890 CAG-----AGTCTGGGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1940
Qy 1991 CAGAGCCCATGTCAGAGATGCTCTGCTGCTGTGTCTCATCTCTGTGTGTGTGTGTGTGTGT 2050
Db 1941 CAGAGAGAGATTAACCAATGCTCTGTGAGCCCAAGGTGTCTGTGCTGTGAGAGCTT 2000
Qy 2051 TCTGTGATGCTTTTCCAAAGCTGTAACTCACTGCTGT 2088
Db 2001 CTCTGGGTGTCTTAAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2038

RESULT 12

AAK67051
ID AAK67051 standard; DNA; 30626 BP.
XX
AC AAK67051;
XX

DT 06-NOV-2001 (first entry)
XX
Db Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21863.
XX
Xw Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
Xw Cytosolic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
PD 09-AUG-2001.
XX
PP 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218293P.
PR 26-JUL-2000; 2000US-0230963P.
PR 26-JUL-2000; 2000US-0230964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0237009P.
PR 30-AUG-2000; 2000US-0238924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229533P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-025190P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 21863; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 30626 BP; 7102 A; 7610 C; 7817 G; 8097 T; 0 U; 0 Other;
Query Match 36.3%; Score 948.4; DB 4; Length 30626;
Best Local Similarity 99.6%; Pred. No. 3.7e-244;
Matches 972; Conservative 0; Mismatches 1; Indels 3; Gaps 2;
QY 1636 CAGGCAAGTATTAATCTCTTCTGCAAGAGACGCTCCAAAAGAGAGACC 1695
DB 4785 CAGGCAAGTATTAATCTCTTCTGCAAGAGACGCTCCAAAAGAGAGACC 4844
QY 1696 ACAGTGAAGAGGCAAGGCTGAGTCCCTGCGGAGGAGACATCGGAGAGAGACGT 1755
DB 4845 ACAGTGAAGAGGCAAGGCTGAGTCCCTGCGGAGGAGACATCGGAGAGAGACGT 4904
QY 1756 TGTGAGAGAGCAAGCGCTGCTGAGGAGAGATCAATTAATCAATCAAGAGAGAG 1815
DB 4905 TGTGAGAGAGCAAGCGCTGCTGAGGAGAGATCAATTAATCAATCAAGAGAGAG 4964
QY 1816 AAGACTGCTGCTCCCTGCGGAGAGATCAATTAATCAATCAAGAGAGAG 1875
DB 4965 AAGACTGCTGCTCCCTGCGGAGAGATCAATTAATCAATCAAGAGAGAG 5024
QY 1876 CTCCTGAGACCCCTCCCTCTTCTGAGATGACATCCCTGAGGCTGTCAGAGATCGGA 1935
DB 5025 CTCCTGAGACCCCTCCCTCTTCTGAGATGACATCCCTGAGGCTGTCAGAGATCGGA 5084
QY 1936 GTTGCAAGCTTGTGTAAGCTGATCAAGACACCGGCTCACCATCAAGAGAGAGAG 1995
DB 5085 GTTGCAAGCTTGTGTAAGCTGATCAAGACACCGGCTCACCATCAAGAGAGAGAG 5144
QY 1996 CCATGTCAGAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2055
DB 5145 CCATGTCAGAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5204
QY 2056 TCAGCTTTTCAAGAGCTGAACCTCACTGGAGAGCTTCACTTAATGATGATTTCT 2115
DB 5205 TCAGCTTTTCAAGAGCTGAACCTCACTGGAGAGCTTCACTTAATGATGATTTCT 5264

QY 2116 TAATCTGTTTCACTCTGAGGCTCTGTAGATTTGATTTCTTCTCATCCAGTCTG 2175
 DB 5265 TAACTCTGTTTCACTCTGAGGCTCTGTAGATTTGATTTCTTCTCATCCAGTCTG 5324
 QY 2176 ATTGCATAGCCACACTGCTCCGCGACGCCACATCCACCCCTGTCTGACATGATGTTCT 2235
 DB 5325 ATTGCATAGCCACACTGCTCCGCGACGCCACATCCACCCCTGTCTGACATGATGTTCT 5384
 QY 2236 GACAAAGCGCTGATATAGCTTCACTTTTTCACATTTCTCCACCGCGACGACATGAAGC 2295
 DB 5385 GACAAAGCGCTGATATAGCTTCACTTTTTCACATTTCTCCACCGCGACGACATGAAGC 5444
 QY 2296 ATCACTCTTTTATGTTGAGGAATCTTTGCAAGTTGATGATGATTTTTCAGG 2355
 DB 5445 ATCACTCTTTTATGTTGAGGAATCTTTGCAAGTTGATGATGATTTTTCAGG 5504
 QY 2356 TGTACATTTATTTTGAATGCTGGGAGATGAGGATTTTTCATGTCGATTTCA 2415
 DB 5505 TGTACATTTATTTTGAATGCTGGGAGATGAGGATTTTTCATGTCGATTTCA 5562
 QY 2416 CACCTTACACACCCACATGAACATTTGAACTTGAAGG--CACACCTCTCTTCACTA 2474
 DB 5563 CACCTTACACACCCACATGAACATTTGAACTTGAAGG--CACACCTCTCTTCACTA 5622
 QY 2475 GAGCCCAACGTAAGTGAAGTCAACCTTAGAACACTGTCTGACCGACGAGCGCTGCTT 2534
 DB 5623 GAGCCCAACGTAAGTGAAGTCAACCTTAGAACACTGTCTGACCGACGAGCGCTGCTT 5682
 QY 2535 GAGCTTGTATTTCTACATGTAATGCTGCTTCTGCTGCTCTTGAATGTTAGACTC 2594
 DB 5683 GAGCTTGTATTTCTACATGTAATGCTGCTTCTGCTGCTCTTGAATGTTAGACTC 5742
 QY 2595 TTAGATCATATCTCTG 2610
 DB 5743 TTAGATCATATCTCTG 5758

RESULT 13

AAS86575
 ID AAS86575 standard; cDNA, 2986 BP.

AC AAS86575;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #22379.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

FN W0200175067-A2.

PD 11-OCT-2001.

PP 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG22388.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 1; SEQ ID NO 22379; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (II) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful for medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

Sequence 2986 BP; 798 A; 796 C; 781 G; 611 T; 0 U; 0 Other;

Query Match 26.2%; Score 682.6; DB 5; Length 2986;

Best Local Similarity 88.1%; Pred. No. 5.3e-173; Indels 116; Gaps 9;

Matches 891; Conservative 0; Mismatches 4;

QY 965 GAGAGAGTAAAGTTGAGGAAAGGCGCTCAGAGATCAAGATGAGAGCTGAGACT 1024

DB 1 GAGAGAGTAAAGTTGAGGAAAGGCGCTCAGAGATCAAGATGAGAGCTGAGACT 60

QY 1025 TGGGCGACCCGAGGCTGAATGTAAGGTTAGACCCGAGCGCTGATTTGCTTGGC 1084

DB 61 TGGGCGACCCGAGGCTGAATGTAAGGTTAGACCCGAGCGCTGATTTGCTTGGC 120

QY 1085 GCCAGCTCACAACAAGGTTTATGATTTGTGTGACCGCTGTGAAGATGTTTCATG 1144

DB 121 GCCAGCTCACAACAAGGTTTATGATTTGTGTGACCGCTGTGAAGATGTTTCATG 180

QY 1145 GCGATTTGTGGGCAATTTCTAGAGCTCGAGGAGGCTTTTGAAGAAAGATGGGAAAGCT 1204

DB 181 GCGATTTGTGGGCAATTTCTAGAGCTCGAGGAGGCTTTTGAAGAAAGATGGGAAAGCT 240

QY 1205 ATATCTGCCAACTGCACTTCTGCAAGTGCAGAGATGAGACTATTCAAGAAAGCGCG 1264

DB 241 ATATCTGCCAACTGCACTTCTGCAAGTGCAGAGATGAGACTATTCAAGAAAGCGCG 300

QY 1265 ATCAGCAGGAAGCTAAATGAGAGCTTGAGATGCTGATGCGACCGATTTGTAAGAATAG 1324

DB 301 ATCAGCAGGAAGCTAAATGAGAGCTTGAGATGCTGATGCGACCGATTTGTAAGAATAG 360

QY 1325 GAAACATAGAGCAGAGAGTCTAGCGAAGACCAAGGATTAAGGTTGAAGAAAGCTG 1384

DB 361 GAAACATAGAGCAGAGAGTCTAGCGAAGACCAAGGATTAAGGTTGAAGAAAGCTG 420

QY 1385 CAATTCAGTGGCAGAGAAAGTCAAGATCTTCCAGCT----- 1425

DB 421 CAATTCAGTGGCAGAGAAAGTCAAGATCTTCCAGCT----- 480

QY 1426 ----- 1425

DB 481 CCCAGCTGCTGTCTCTGCGAGGATTTGAAATTTGCTGTGCTAGAGCATCTGCT 540

QY 1426 -----GTGATAGAGCGCTGTGCTCAAAATGTA 1456

DB 541 TCACACTTTGACATGATTAAGCTCAAGATGATAGAGCGCTGTGCTCAAAATGTA 600

QY 1457 TTGGCCCC--GGGTGCTGTCACTGGCGCAG--CCGCACTGGGTGCTGCA--GTAAATGACT 1513

DB 601 TTGGCCCCGGGGTGTGTGACGTGGCGAGCCCGCACTGGGTGCTGCA--GTAAATGACT 660

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:42:12 ; Search time 9937.73 Seconds
(without alignments)
11383.418 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 2610
Sequence: 1 ctgcgtgscgcgtccgcacac.....actctaagatcatatccctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBml.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_rfc.*
12: gb_ey.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pac.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_str.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_in.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pin.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hngo_hum.*
40: em_hngo_mus.*
41: em_hngo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB	ID	Description
1	2610	100.0	2610	6	BD244712	BD244712 Genes enc
2	2610	100.0	2610	6	AX023369	AX023369 Sequence
3	2532.4	97.0	2725	9	BC014489	BC014489 Homo sapi
4	2517.8	96.5	2767	9	BC004237	BC004237 Homo sapi
5	2516.2	96.4	2772	6	AX878224	AX878224 Sequence
6	2516.2	96.4	2772	6	BD157094	BD157094 Primer fo
7	2516.2	96.4	2772	9	AK002127	AK002127 Homo sapi
8	1941	74.4	2407	9	BC000770	BC000770 Homo sapi
9	1023.6	39.2	2602	10	MM0238332	AJ38332 Mus muscu
10	1023.6	39.2	2867	6	BD244713	BD244713 Genes enc
11	1023.6	39.2	2867	6	AX023370	AX023370 Sequence
12	985	37.7	4906	10	AX129117	AX129117 Mus muscu
13	948.4	36.3	160241	9	HS885L7	AL035669 Human DNA
14	770.6	29.5	6692	9	AB002331	AB002331 Human mRN
15	654.6	25.1	28519	9	HS7563B14	AL117379 Human DNA
16	596.8	22.9	659	6	AX897185	AX897185 Sequence
17	596.8	22.9	659	6	BD032718	BD032718 Sequence
18	510.8	19.6	681	6	AX869935	AX869935 Sequence
19	510.8	19.6	681	6	BD149997	BD149997 Primer fo
20	491.4	18.8	650	6	HMMY10D03	AF085876 Homo sapi
21	490.4	18.8	272404	2	AC108337	AC108337 Rattus no
22	481.8	18.5	104653	10	AL732560	AL732560 Mouse DNA
23	449.8	17.2	221547	2	AC119716	AC119716 Rattus no
24	386.4	14.8	69252	2	AC101519	AC101519 Mus muscu
25	386.4	12.9	542	6	AX874900	AX874900 Sequence
26	336.6	12.9	542	6	BD154962	BD154962 Primer fo
27	262.4	10.1	2428	5	BC060442	BC060442 Xenopus 1
28	232	8.9	69252	2	AC101519	AC101519 Mus muscu
29	202.8	7.8	472	6	AR426988	AR426988 Sequence
30	202.8	7.8	472	6	BD122541	BD122541 EST and e
31	103.4	4.0	198133	3	BX296530	BX296530 Dario rer
32	70.8	2.7	2161	2	AK114522	AK114522 Ciona int
33	68.6	2.6	61204	2	AC017132	AC017132 Drosophi1
34	68.6	2.6	161601	3	AC007594	AC007594 Drosophi1
35	68.6	2.6	225655	3	AB003695	AB003695 Drosophi1
36	67	2.6	7218	6	166494	166494 Sequence 14
37	66.2	2.5	125020	9	AF429315	AF429315 Homo sapi
38	65	2.5	2888	9	AK055678	AK055678 Homo sapi
39	65	2.5	5975	6	AX329603	AX329603 Sequence
40	65	2.5	6256	6	AR338778	AR338778 Sequence
41	65	2.5	6936	9	D87685	D87685 Human mRNA
42	65	2.5	6948	6	AX210673	AX210673 Sequence
43	65	2.5	6948	9	AF091622	AF091622 Homo sapi
44	65	2.5	7142	9	HSW808416	BX648268 Homo sapi
45	64	2.5	125020	9	AF429315	AF429315 Homo sapi

ALIGNMENTS

RESULT 1
BD244712 2610 bp DNA linear PAT 17-JUL-2003
LOCUS BD244712
DEFINITION BD244712 Genes encoding for the human and murine death
inducer-obliterato-1.
ACCESSION BD244712
VERSION BD244712.1 GI:33054482
KEYWORDS JP 2002526040-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2610)
Alonso, C.M., Domingo, D.G., Grandien, A., Leonardo, E. and Martinez, P.
TITLE Genes encoding for the human and murine death inducer-obliterato-1

QY	1621	CCGAATATCGGTGCTCAGCCAGAGATTAAATTCCTTCCTGTCGACAAGAGCCAGCTCCA	1680
Db	1621	CCGAATATCGGTGCTCAGCCAGAGATTAAATTCCTTCCTGTCGACAAGAGCCAGCTCCA	1680
QY	1681	GAATAAAGAGAGACCAAGTGAAGAAGGACAGTGTGTCCTCGCGGAGTGAAGACATC	1740
Db	1681	GAATAAAGAGAGACCAAGTGAAGAAGGACAGTGTGTCCTCGCGGAGTGAAGACATC	1740
QY	1741	GGAGAGAGACGCTTGTGAGAGACACCGCTGTGTGTGTGTGTAAATGTATGAT	1800
Db	1741	GGAGAGAGACGCTTGTGAGAGACACCGCTGTGTGTGTGTGTAAATGTATGAT	1800
QY	1801	GCATTAAGCCAGAAAAGACATGCTGCTCCCTCGCGGTCACTGTGTATTAATGTATGAT	1860
Db	1801	GCATTAAGCCAGAAAAGACATGCTGCTCCCTCGCGGTCACTGTGTATTAATGTATGAT	1860
QY	1861	CACCTAAGGGGTGGCTCTCTGGAACCCCTCCCGTTCCTTCTGGAATAGCCATCCCTGGGCC	1920
Db	1861	CACCTAAGGGGTGGCTCTCTGGAACCCCTCCCGTTCCTTCTGGAATAGCCATCCCTGGGCC	1920
QY	1921	TGTCAGAGACTGGAGATTGCAGCTTTGTGTAACTGATTCACAGACACCGGCTGCACAT	1980
Db	1921	TGTCAGAGACTGGAGATTGCAGCTTTGTGTAACTGATTCACAGACACCGGCTGCACAT	1980
QY	1981	CAGCGGGAAGGAGAGCCCATGTCCAGAGATGCTCTGTGCTCCGTGTGCATCCCTAGTC	2040
Db	1981	CAGCGGGAAGGAGAGCCCATGTCCAGAGATGCTCTGTGCTCCGTGTGCATCCCTAGTC	2040
QY	2041	TGTCAGAGACTCTGTCACTGTGTTTCCAAAGCTGTAAACCTCACTGTGTGAAGTTCACT	2100
Db	2041	TGTCAGAGACTCTGTCACTGTGTTTCCAAAGCTGTAAACCTCACTGTGTGAAGTTCACT	2100
QY	2101	TAAATGATTGATTTCTTTAAATCTCTGTGTTTCACTTCAGGCTCTGTGAATATTGATTC	2160
Db	2101	TAAATGATTGATTTCTTTAAATCTCTGTGTTTCACTTCAGGCTCTGTGAATATTGATTC	2160
QY	2161	CTTCATCCCAAGTCTGATTTGATTAACCACTGCTCCGACGCCCATCTCACCCCTGTCTG	2220
Db	2161	CTTCATCCCAAGTCTGATTTGATTAACCACTGCTCCGACGCCCATCTCACCCCTGTCTG	2220
QY	2221	CACATGAGTTGTCTGACAAAGCGCTGTAAATACGTTTCAGTTTTCACATTTGTCACGG	2280
Db	2221	CACATGAGTTGTCTGACAAAGCGCTGTAAATACGTTTCAGTTTTCACATTTGTCACGG	2280
QY	2281	CCAGCACAATGAACACATCACTTCTTTTAAAGTTGTGGGAATCTTGGAAATTAAGTGTG	2340
Db	2281	CCAGCACAATGAACACATCACTTCTTTTAAAGTTGTGGGAATCTTGGAAATTAAGTGTG	2340
QY	2341	CATCTGATTTTCAAGTGTACATTTATTTTGAATGAGGAGATTTTATTTT	2400
Db	2341	CATCTGATTTTCAAGTGTACATTTATTTTGAATGAGGAGATTTTATTTT	2400
QY	2401	TCCATGTCCGATTCACACGCTTACACCCACATGAACATTTGCAATTTGCAAGGACATC	2460
Db	2401	TCCATGTCCGATTCACACGCTTACACCCACATGAACATTTGCAATTTGCAAGGACATC	2460
QY	2461	ACTCTGTCTTATAGGCCCAACGGTAAATGAGTTCACTAGGAACATGTCTGTACCGC	2520
Db	2461	ACTCTGTCTTATAGGCCCAACGGTAAATGAGTTCACTAGGAACATGTCTGTACCGC	2520
QY	2521	AGAGACGATGCTTGAATTTGATTTTCAATGATGATCTGCTTCTGCTCTGCTCTCT	2580
Db	2521	AGAGACGATGCTTGAATTTGATTTTCAATGATGATCTGCTTCTGCTCTGCTCTCT	2580
QY	2581	GAATGTTTAACTCTTAAAGATCAATATCTG	2610
Db	2581	GAATGTTTAACTCTTAAAGATCAATATCTG	2610

RESULT 2			
AX023369			
LOCUS	AX023369	2610 bp	DNA
DEFINITION	Sequence 1 from Patent WO0015787.	linear	PAT 15-SBP-2000

ACCESSION	AX023369
VERSION	AX023369.1
KEYWORDS	GI:10183781
SOURCE	
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	
JOURNAL	
FEATURES	
source	
1	CGCGTGGCCGCTCCGCGCCCACTCGGCGCGCTTGGGGAATAGGCTGGAGACCTTAGAGGC
2	
3	
4	1 CTGGTGGCGCTCCGCGCCCACTCGGCGCGCTTGGGGAATAGGCTGGAGACCTTAGAGGC
5	
6	
7	61 CTGGGAGCTTACTCCACGGGAAACAGGCTTAGATATCTGAATGTTGTAATACGAG
8	
9	61 CTGGGAGCTTACTCCACGGGAAACAGGCTTAGATATCTGAATGTTGTAATACGAG
10	
11	61 CTGGGAGCTTACTCCACGGGAAACAGGCTTAGATATCTGAATGTTGTAATACGAG
12	
13	121 CCGTGTACTCGTGAACAGTGGCTGACCAACAGTGTGTTGTGAGCTCGGCTGTCTTGG
14	
15	121 CCGTGTACTCGTGAACAGTGGCTGACCAACAGTGTGTTGTGAGCTCGGCTGTCTTGG
16	
17	121 CCGTGTACTCGTGAACAGTGGCTGACCAACAGTGTGTTGTGAGCTCGGCTGTCTTGG
18	
19	181 ACCCAGAGGTTTGTCTGCGCAGGGTTTTTGGTGTATTAGAGATTTCAGGAAAGTCT
20	
21	181 ACCCAGAGGTTTGTCTGCGCAGGGTTTTTGGTGTATTAGAGATTTCAGGAAAGTCT
22	
23	181 ACCCAGAGGTTTGTCTGCGCAGGGTTTTTGGTGTATTAGAGATTTCAGGAAAGTCT
24	
25	241 CAAGCTTTCAGTGTGAGCAGGATATGAGACAAAGGCGACCCGACCAATGAGAGGCA
26	
27	241 CAAGCTTTCAGTGTGAGCAGGATATGAGACAAAGGCGACCCGACCAATGAGAGGCA
28	
29	241 CAAGCTTTCAGTGTGAGCAGGATATGAGACAAAGGCGACCCGACCAATGAGAGGCA
30	
31	301 CCTAAGGCATCAAAACCCACACAGCAAAAGATTCCAGAAAAATGGGGTTTTCCAGGACC
32	
33	301 CCTAAGGCATCAAAACCCACACAGCAAAAGATTCCAGAAAAATGGGGTTTTCCAGGACC
34	
35	301 CCTAAGGCATCAAAACCCACACAGCAAAAGATTCCAGAAAAATGGGGTTTTCCAGGACC
36	
37	361 ACTATGCCCAAGGAGAGGCGCGCAGGAGACCGGAGGCTTGAACCTATGAGGCCCAACC
38	
39	361 ACTATGCCCAAGGAGAGGCGCGCAGGAGACCGGAGGCTTGAACCTATGAGGCCCAACC
40	
41	361 ACTATGCCCAAGGAGAGGCGCGCAGGAGACCGGAGGCTTGAACCTATGAGGCCCAACC
42	
43	421 CCAACAGAGAGAGTGGGCGCTGTCCCTGCGCGGCGAGTGGAGGGCAGCCCAAGCCGACTGAG
44	
45	421 CCAACAGAGAGAGTGGGCGCTGTCCCTGCGCGGCGAGTGGAGGGCAGCCCAAGCCGACTGAG
46	
47	421 CCAACAGAGAGAGTGGGCGCTGTCCCTGCGCGGCGAGTGGAGGGCAGCCCAAGCCGACTGAG
48	
49	481 CGCGTGGAGAGATTCTGAGCAATTGGCGCGCGCGCGAGGAGAGCATGGCTGTCTCC
50	
51	481 CGCGTGGAGAGATTCTGAGCAATTGGCGCGCGCGCGAGGAGAGCATGGCTGTCTCC
52	
53	481 CGCGTGGAGAGATTCTGAGCAATTGGCGCGCGCGCGAGGAGAGCATGGCTGTCTCC
54	
55	541 CTGAGAGATTCTGAGAGCCCAAGTCTGCGCCGACCAAGCCGCGAGACAGCTCCGAG
56	
57	541 CTGAGAGATTCTGAGAGCCCAAGTCTGCGCCGACCAAGCCGCGAGACAGCTCCGAG
58	
59	541 CTGAGAGATTCTGAGAGCCCAAGTCTGCGCCGACCAAGCCGCGAGACAGCTCCGAG
60	
61	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
62	
63	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
64	
65	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
66	
67	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
68	
69	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
70	
71	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
72	
73	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
74	
75	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
76	
77	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
78	
79	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
80	
81	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
82	
83	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
84	
85	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
86	
87	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
88	
89	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
90	
91	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
92	
93	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
94	
95	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
96	
97	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
98	
99	601 GCGAGCGCTGAAAAGCGCTTCTGAGACCAAGAGGGGCCCGCAGTGTCTTCCACAGCTGTG
100	

```

Qy 661 AAGGAAGACAGCCTCTTCTGAAAAGGTGAAGAGGGGATGACCAAGATGACACCTCC 720
Db 661 AAGGAAGACAGCCTCTTCTGAAAAGGTGAAGAGGGGATGACCAAGATGACACCTCC 720
Qy 721 GATAGTACAGCAGTGGCTGACCTTGAAGAGCTTCAGAAATGCGCTTCGAGGAAGCGG 780
Db 721 GATAGTACAGCAGTGGCTGACCTTGAAGAGCTTCAGAAATGCGCTTCGAGGAAGCGG 780
Qy 781 GAACAGAGACCCATCTGAGAGGCCCTTGAAGAGGATCCAGATCCGCTGCGGAAGAGCGC 840
Db 781 GAACAGAGACCCATCTGAGAGGCCCTTGAAGAGGATCCAGATCCGCTGCGGAAGAGCGC 840
Qy 841 CGGAGAGAGGGTCCCGCGAGACTGTGGGCTCCAGAGCCAGTGAACATGTGAGAGGGCTC 900
Db 841 CGGAGAGAGGGTCCCGCGAGACTGTGGGCTCCAGAGCCAGTGAACATGTGAGAGGGCTC 900
Qy 901 CTGCCAGTAAAGCAGAGAGCCCGAGAGCAGATCAGGGGGTGTGTCTCCAGAGCTGGGAAGAT 960
Db 901 CTGCCAGTAAAGCAGAGAGCCCGAGAGCAGATCAGGGGGTGTGTCTCCAGAGCTGGGAAGAT 960
Qy 961 GACAGAGAGTAAAGTGTGAGAGGAAAGCGGCTCAGACATCAAGATGAGAGAGCTGGA 1020
Db 961 GACAGAGAGTAAAGTGTGAGAGGAAAGCGGCTCAGACATCAAGATGAGAGAGCTGGA 1020
Qy 1021 GACTTGGGGCGAGCGAAGCCGTAAGTGTAGGGGTTACGACCCCAAGCCCTGTATTGCAAT 1080
Db 1021 GACTTGGGGCGAGCGAAGCCGTAAGTGTAGGGGTTACGACCCCAAGCCCTGTATTGCAAT 1080
Qy 1081 TGCCGCGAGCTCTCAACAACAAGGTTTATGATTTGCTGTGACCGCTGTGAAGATGTTTT 1140
Db 1081 TGCCGCGAGCTCTCAACAACAAGGTTTATGATTTGCTGTGACCGCTGTGAAGATGTTTT 1140
Qy 1141 CATGGCGATTGTGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGAAGAGATGGGGA 1200
Db 1141 CATGGCGATTGTGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGAAGAGATGGGGA 1200
Qy 1201 GACTATATCTGCGCAAACTGCAACATCTGCAAGTGTGAGAGTGAAGATCAATTCAGAAAG 1260
Db 1201 GACTATATCTGCGCAAACTGCAACATCTGCAAGTGTGAGAGTGAAGATCAATTCAGAAAG 1260
Qy 1261 GCAGATCAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGCAGATTTGACAGT 1320
Db 1261 GCAGATCAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGCAGATTTGACAGT 1320
Qy 1321 ATAGGAACAATGAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGAA 1380
Db 1321 ATAGGAACAATGAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGAA 1380
Qy 1381 GCTGCAATTCAGAGTGTGAGAGAGAACTCAAGATCTTCAAGCTGTGATGAGAGCGCT 1440
Db 1381 GCTGCAATTCAGAGTGTGAGAGAGAACTCAAGATCTTCAAGCTGTGATGAGAGCGCT 1440
Qy 1441 GGTGCTCTCAAAATGTATTTGGCCCGGGTGTCTGACGTGGCGCAGCCGACTCGTGTAC 1500
Db 1441 GGTGCTCTCAAAATGTATTTGGCCCGGGTGTCTGACGTGGCGCAGCCGACTCGTGTAC 1500
Qy 1501 TGCAGTATGACTGTATCTTCAACACCGCCGAGAGCAATGAAATTTCTAAGCTCAAGT 1560
Db 1501 TGCAGTATGACTGTATCTTCAACACCGCCGAGAGCAATGAAATTTCTAAGCTCAAGT 1560
Qy 1561 AAAGGAAGAGAGCCAAAGCCTTAAAGAAAGTGAAGATGAAGGCCAGAGAGCCAGTCTT 1620
Db 1561 AAAGGAAGAGAGCCAAAGCCTTAAAGAAAGTGAAGATGAAGGCCAGAGAGCCAGTCTT 1620
Qy 1621 CCGAATATCGGTGCTCAGCAGAGTAAATCTTCTGTGACACAAGAGCCAGCTCCA 1680
Db 1621 CCGAATATCGGTGCTCAGCAGAGTAAATCTTCTGTGACACAAGAGCCAGCTCCA 1680
Qy 1681 GAAAAAAGAGAGCACAAGTGAAGAGAGCAGTGTGTCTCCGCGAGTGAAGACACTC 1740
Db 1681 GAAAAAAGAGAGCACAAGTGAAGAGAGCAGTGTGTCTCCGCGAGTGAAGACACTC 1740

```

```

Qy 1741 GGAAGGAAGCAGCTGTGTGAGAGAGCAGCAGCCGCTGTGTGGGAGAGATCAATTAAT 1800
Db 1741 GGAAGGAAGCAGCTGTGTGAGAGAGCAGCAGCCGCTGTGTGGGAGAGATCAATTAAT 1800
Qy 1801 GCAGTAAAGCCAGAAAAGACTGTGTCTGCTCCCTGCTCAGTGTGTATTAATGTAT 1860
Db 1801 GCAGTAAAGCCAGAAAAGACTGTGTCTGCTCCCTGCTCAGTGTGTATTAATGTAT 1860
Qy 1861 CACCTAGGGGTGGCTCTCCAGAGCCCGCTCCCTTCTTCTGAGATAGCATCCCTGGGCT 1920
Db 1861 CACCTAGGGGTGGCTCTCCAGAGCCCGCTCCCTTCTTCTGAGATAGCATCCCTGGGCT 1920
Qy 1921 TGTCCAGAGCTGGAGGTGACGCTTGTGTGATGATGATCAAGACACCGCTGACCAT 1980
Db 1921 TGTCCAGAGCTGGAGGTGATGAGCTTGTGTGATGATGATGATCAAGACACCGCTGACCAT 1980
Qy 1981 CAGCGGAGAGCAGAGCCCATATGTCAGATGCTCTGCTGCTGCTGTGTCATTCCTATCT 2040
Db 1981 CAGCGGAGAGCAGAGCCCATATGTCAGATGCTCTGCTGCTGCTGTGTCATTCCTATCT 2040
Qy 2041 TGTCCAGAGCTTCCGTGACGCTTTTCCAAAGCTGTAAACCTCACTGTGAAGTTCACCT 2100
Db 2041 TGTCCAGAGCTTCCGTGACGCTTTTCCAAAGCTGTAAACCTCACTGTGAAGTTCACCT 2100
Qy 2101 TAATGATGATTTCTTAAATCTCTGTTTTCACTCAGGCTCTGATGATTTGATTTCT 2160
Db 2101 TAATGATGATTTCTTAAATCTCTGTTTTCACTCAGGCTCTGATGATTTGATTTCT 2160
Qy 2161 CTTATATCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
Db 2161 CTTATATCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
Qy 2221 CACATGAGTGTGTGACAAAGAGGCTGTATAGCTTCAAGTTTTCACATTTGACAGG 2280
Db 2221 CACATGAGTGTGTGACAAAGAGGCTGTATAGCTTCAAGTTTTCACATTTGACAGG 2280
Qy 2281 CCAGACATGAAAGCATCTCTTTTATATGTGTGTGAGAACTTTGCAAGTATGATTTG 2340
Db 2281 CCAGACATGAAAGCATCTCTTTTATATGTGTGTGAGAACTTTGCAAGTATGATTTG 2340
Qy 2341 CATCTGATTTTCAAGTGTACATTTATTTTGTGATGTGGAGATAGGGATTTTGT 2400
Db 2341 CATCTGATTTTCAAGTGTACATTTATTTTGTGATGTGGAGATAGGGATTTTGT 2400
Qy 2401 TCCATGTCGATTCACAGCTACACGCTACACCACTAATGACATTCCTGAAAGCAGC 2460
Db 2401 TCCATGTCGATTCACAGCTACACGCTACACCACTAATGACATTCCTGAAAGCAGC 2460
Qy 2461 ACTCTGCTTCAATAGGCCCAAGGTATGATGATGATGATGATGATGATGATGATGAT 2520
Db 2461 ACTCTGCTTCAATAGGCCCAAGGTATGATGATGATGATGATGATGATGATGATGAT 2520
Qy 2521 AGAAGCGCTGCTTGGACTTGGATTTCTACATGTGACGTGTTTGTGCTGCTCTT 2580
Db 2521 AGAAGCGCTGCTTGGACTTGGATTTCTACATGTGACGTGTTTGTGCTGCTCTT 2580
Qy 2581 GAATGTTAGACTTTAAGATCATATCTG 2610
Db 2581 GAATGTTAGACTTTAAGATCATATCTG 2610

```

```

RESULT 3
BC014489 2725 bp mRNA linear PRI 04-OCT-2003
LOCUS BC014489
DEFINITION Homo sapiens deach associated transcription factor 1, complete cds.
ACCESSION BC014489
VERSION BC014489.1 GI:15680266
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```


QY 766 CTTGCGAAGAACCGGGAACAGAGCCCACTGAGAGGCCCTCTGAAGAGATCCAGAGTGC 825
 DB 739 CTTGCGAAGAACCGGGAACAGAGCCCACTGAGAGGCCCTCTGAAGAGATCCAGAGTGC 798
 QY 826 CTTGCGAAGAACCGGGAAGAGAGGCTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTAC 885
 DB 799 CTTGCGAAGAACCGGGAAGAGAGGCTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTAC 858
 QY 886 ACTGTGAGGGGCTCTCCCGAGTAAAGAGAGGCCGAGAACGATCAGGGGGTGTGTCC 945
 DB 859 ACTGTGAGGGGCTCTCCCGAGTAAAGAGAGGCCGAGAACGATCAGGGGGTGTGTCC 918
 QY 946 CAGGCTGGGAAAGAGTAAAGAGAGAGTAAAGTGGAGGAAAGCCGCTCAGGACATCAAA 1005
 DB 919 CAGGCTGGGAAAGAGTAAAGAGAGAGTAAAGTGGAGGAAAGCCGCTCAGGACATCAAA 978
 QY 1006 GATGAGAGCCCTGAGAGCTTGGGCGCAGCCGAGCCCTGAATGTAGAGGTTACGACCCCAAC 1065
 DB 979 GATGAGAGCCCTGAGAGCTTGGGCGCAGCCGAGCCCTGAATGTAGAGGTTACGACCCCAAC 1038
 QY 1066 GCCCTGTATTTGATTTGCCCGCAGCTTCAACAAACAGGTTTATGATTTGCTGTACCGC 1125
 DB 1039 GCCCTGTATTTGATTTGCCCGCAGCTTCAACAAACAGGTTTATGATTTGCTGTACCGC 1098
 QY 1126 TGTGAGAAATGGTTTCAATGCGAGTGTGTGGGCTTTCTGAGGCTCGAGGGAGCTTTTG 1185
 DB 1099 TGTGAGAAATGGTTTCAATGCGAGTGTGTGGGCTTTCTGAGGCTCGAGGGAGCTTTTG 1158
 QY 1186 GAAAGAGATGGGGAAGACTATATCTGCCCAACCTGACCATTTCTGCAAGTGCAGAGTAG 1245
 DB 1159 GAAAGAGATGGGGAAGACTATATCTGCCCAACCTGACCATTTCTGCAAGTGCAGAGTAG 1218
 QY 1246 ACTCATTCAGAAACCGCAGATCAGAGAGAGTAAATGAGACCTGAGAGTCTGATGCGC 1305
 DB 1219 ACTCATTCAGAAACCGCAGATCAGAGAGAGTAAATGAGAGCTGAGAGTCTGATGCGC 1278
 QY 1306 ACCGATTTACAGATGATGAGAACATAGACAGAGTCTGAGGAGAGACCAAGGATTAAG 1365
 DB 1279 ACCGATTTACAGATGATGAGAACATAGAGAGAGTCTGAGGAGAGACCAAGGATTAAG 1338
 QY 1366 GGTAGAAATTTGAGAAAGCTGCAATCTCAAGTGCAGAGAAAGCTCAAGATCTTCCAGCT 1425
 DB 1339 GGTAGAAATTTGAGAAAGCTGCAATCTCAAGTGCAGAGAAAGCTCAAGATCTTCCAGCT 1398
 QY 1426 GTGATAGAGGCGCTGTGTCTCAAAATGATTTGGCCCGGTGCTGTACGTTGCGCAG 1485
 DB 1399 GTGATAGAGGCGCTGTGTCTCAAAATGATTTGGCCCGGTGCTGTACGTTGCGCAG 1458
 QY 1486 CCGGACTGGGTGTACTGCGAGTAAATGATCTTCAAAACAGCCGCGACGCAATTAAG 1545
 DB 1459 CCGGACTGGGTGTACTGCGAGTAAATGATCTTCAAAACAGCCGCGACGCAATTAAG 1518
 QY 1546 TTTCTAAGCTCAGGTAAAGAACAGAACCCAAAGCTTAAAGAAAGTGAAGTGAAGCCA 1605
 DB 1519 TTTCTAAGCTCAGGTAAAGAACAGAACCCAAAGCTTAAAGAAAGTGAAGTGAAGCCA 1578
 QY 1606 GAGAGCCCAAGTCTTCCGAATGCGGTGCTCAGGCGAGTATTAATCTTCTGTGCAAC 1665
 DB 1579 GAGAGCCCAAGTCTTCCGAATGCGGTGCTCAGGCGAGTATTAATCTTCTGTGCAAC 1638
 QY 1666 AAGAGACCAAGCTCCGAAAGAAAGAAAGACACAGAGTAAAGAGAGAGTGTGCTCCCTG 1725
 DB 1639 AAGAGACCAAGCTCCGAAAGAAAGAAAGACACAGAGTAAAGAGAGAGTGTGCTCCCTG 1698
 QY 1726 CGGAGTGAAGCACTCGGGAAGAGAGAGCTTGTGAGAGAGACGCGCTGTGTGCGAGAC 1785
 DB 1699 CGGAGTGAAGCACTCGGGAAGAGAGAGCTTGTGAGAGAGACGCGCTGTGTGCGAGAC 1758
 QY 1786 GATCAACAATTAATGATGAGTAAAGCCAGAAAGACTGTGCTCTCCGCTGACATGTTG 1845
 DB 1759 GATCAACAATTAATGATGAGTAAAGCCAGAAAGACTGTGCTCTCCGCTGACATGTTG 1818
 QY 1846 TATAAATGATATACCTACCTAGGAGGCTGTGCGACCCCTCCGCTTCTTGTGATATA 1905

DB 1819 TATAAATGATATACCTACCTAGGAGGCTGTGCGACCCCTCCGCTTCTTGTGATATA 1878
 QY 1906 GCCATCCCTCGGCGCTGTCCAGAGATTTGGAAGTGTGATTTGATTAAGTCAACAGA 1965
 DB 1879 GCCATCCCTCGGCGCTGTCCAGAGATTTGGAAGTGTGATTTGATTAAGTCAACAGA 1938
 QY 1966 CACCGGCTGACACATCAGCGGGAACAGAGCCCATGTCCAGAGATGCTCTGCTCCCTG 2025
 DB 1939 CACCGGCTGACACATCAGCGGGAACAGAGCCCATGTCCAGAGATGCTCTGCTCCCTG 1998
 QY 2026 TGTTCATCCCTAGTCTGTGAGAGCTTCTGTCACTGTGTTTCCAAAGCTGTAACCTCACT 2085
 DB 1999 TGTTCATCCCTAGTCTGTGAGAGCTTCTGTCACTGTGTTTCCAAAGCTGTAACCTCACT 2058
 QY 2086 GGTGAACCTTCACTTAATGATTAATTTCTTAAATCTGTGTTTCACTGACGCTGTGT 2145
 DB 2059 GGTGAACCTTCACTTAATGATTAATTTCTTAAATCTGTGTTTCACTGACGCTGTGT 2118
 QY 2146 AAGTATTTGATATCTCTTCATCCAGTCTGATTTGATGAGCCACATGCCCCGACGCCAC 2205
 DB 2119 AAGTATTTGATATCTCTTCATCCAGTCTGATTTGATGAGCCACATGCCCCGACGCCAC 2178
 QY 2206 ATCCACCCCTGTCTGACATGAGTGTCTGACAAACAGGCTGTATPAGCTTCACTTTT 2265
 DB 2179 ATCCACCCCTGTCTGACATGAGTGTCTGACAAACAGGCTGTATPAGCTTCACTTTT 2238
 QY 2266 CCACATTTGTCACGCGCCAGACATGAAAGCATCTTTCTTTTATGTTGTTGGAATCTT 2325
 DB 2239 CCACATTTGTCACGCGCCAGACATGAAAGCATCTTTCTTTTATGTTGTTGGAATCTT 2298
 QY 2326 TGCAGTATGATGTTGATCTGATTTTCAAGTGTATGATTTTGAATGTTGAGTGAAG 2385
 DB 2299 TGCAGTATGATGTTGATCTGATTTTCAAGTGTATGATTTTGAATGTTGAGTGAAG 2358
 QY 2386 GGAATTTTGTGTTTTCATGTCAGATTCACAGCTTACACCCACATGACATTCATGCA 2445
 DB 2359 GGAATTTTGTGTTTTCATGTCAGATTCACAGCTTACACCCACATGACATTCATGCA 2416
 QY 2446 ACTTGGAAGG-CACACATCTCTGCTTCAATGAGCCCAAGTAAAGTGAAGTCAACCTAGA 2504
 DB 2417 ACTTGGAAGG-CACACATCTCTGCTTCAATGAGCCCAAGTAAAGTGAAGTCAACCTAGA 2476
 QY 2505 AACCTGCTGACCGCAGAGAGCGTGCCTTGAATGTTGATTTCTACATGATGATGCTTT 2564
 DB 2477 AACCTGCTGACCGCAGAGAGCGTGCCTTGAATGTTGATTTCTACATGATGATGCTTT 2536
 QY 2565 CTTGCGCTCGTCTCTTGAATGTTTGAATCTTTAAGATCATATCTG 2610
 DB 2537 CTTGCGCTCGTCTCTTGAATGTTTGAATCTTTAAGATCATATCTG 2582

RESULT 4
 BC004237 2767 bp mRNA linear PRI 12-NOV-2003
 LOCUS
 DEFINITION Homo sapiens death associated transcription factor 1, transcript variant 1, mRNA (cdna clone MGC:896 IMAGS:2959493), complete cds.
 ACCESSION BC004237 GI:33872756
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 2767)
 Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ueblin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 19, 2003, this sequence version replaced gi:13278980.
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: JAL Plate: 2 Row: 0 Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18375620.

FEATURES

SOURCE

gene

CDS

misc_feature

```
1. 2767
  /organism="Homo sapiens"
  /mol_type="rRNA"
  /db_xref="taxon:9606"
  /clone="MGC:896 IMAG:2959493"
  /tissue_type="Colon, adenocarcinoma"
  /clone_1lb="NIH MGC 15"
  /lab_host="DH10B-R"
  /note="Vector: pOTB7"
  1.. 2767
  /gene="DAP1"
  /note="synonyms: DIO1, FLJ11265, KIA0333"
  /db_xref="locusid:11083"
  /db_xref="MIM:604140"
  282.. 1970
  /codon_start=1
  /product="death associated transcription factor 1, isoform
  a"
  /protein_id="AAH04237.1"
  /db_xref="gi:13278981"
  /translation="MDDKDPSENEAPRAIKPTSKPRRTWGFRRTTAKREGADAS
  ADPLEPPPOOILGLSRSGRQKRTGRBOFLIARRRGRSMPSVLSDEGRTSC
  PATDLETSBSVBSASTRSQPSASTVAKRPSBSERVKGDHDTSDSDSGLT
  LKELQRLRKREKDEPFRPLKGIQSLRKRRRGPATVGSASDVVGLPSKQ
  PENQGVVSQAGKDEPRLRSKGAODIKDEPGLGRPKPCBYDPVALYCIKOP
  HNNEFMICDCEWPHDCVGISARAGLERNEDYICPCTTILQVODETHSTAD
  OQBAKRRGDAGDCTGISTIGTIBOKSSEDOGKGRIBKANPSGKKKXIFQPIVLEAP
  GASKTIGGCCVADPDVSVCSNDCILGHAAATMKPLISGKGRKPKPKMKRKP
  SLPKCAQAGTICISSVHKRPAPKPKETTYKKAIVVVPASBAIGKAKCESSPSSAASD
  HNYNAVKERTPAAPSPILYKCMHVLGVGLDPSRSFWALIFWACPGJGVAAAC"
  1092.. 1241
```

ORIGIN

Query Match 96.5%; Score 2517.8; DB 9; Length 2767;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2542; Conservative 0; Mismatches 2; Indels 3; Gaps 2;

/note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="CDD:pfam00628"

```
65 GGAGCTTACCTCCAGGGGAAAGCCCTGATATCTGAGTTGTGTAATAGAACCTG 124
82 GGAGCTTATTCACGGGAAAGCCCTGATATCTGAGTTGTGTAATAGAACCTG 141
125 TTACTCTGTAACAGTGGCTGACAAAGTGTGTGAGCTGTGCTGTGTTGACCC 184
142 TTACTCTGTAACAGTGGCTGACAAAGTGTGTGAGCTGTGCTGTGTTGACCC 201
185 AGAGGTTTGTCTGTCAGGGTTTGTGTGTAATTTAGATTTCAGGAAAAGTGTCAA 244
202 AGAGGTTTGTCTGTCAGGGTTTGTGTGTAATTTAGATTTCAGGAAAAGTGTCAA 261
245 CTTTCAAGTTTGGACAGATATGACAAAGGGAACCCGACAAATGAGAGGACCTA 304
262 CTTTCAAGTTTGGACAGATATGACAAAGGGAACCCGACAAATGAGAGGACCTA 321
305 AGGACATCAAAACCCACAGCAAAAGATTCAAGAAAACATGAGGTTTTCAGAGCACTA 364
322 AGGACATCAAAACCCACAGCAAAAGATTCAAGAAAACATGAGGTTTTCAGAGCACTA 381
365 TCGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 424
382 TCGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 441
425 AGCAGAGCTGGGCTCTGTCCCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 484
442 AGCAGAGCTGGGCTCTGTCCCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 501
485 TGGAGCAGTCTCTGACATTGCGCGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 544
502 TGGAGCAGTCTCTGACATTGCGCGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 561
545 AGGATTTGTGTGAGGCTTGTGCTGCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 604
562 AGGATTTGTGTGAGGCTTGTGCTGCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 621
605 GCGTGAAGAGGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 664
622 GCGTGAAGAGGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 681
665 AAGGACAGGCTCTTCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 724
682 AAGGACAGGCTCTTCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 741
725 GTGACAGGAGTGGCTGACCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 784
742 GTGACAGGAGTGGCTGACCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 801
785 AAGGAGCCCACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 844
802 AAGGAGCCCACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 861
845 AAGGAGGTCCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 904
862 AAGGAGGTCCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 921
905 CCAAGTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 964
922 CCAAGTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 981
965 GAGAGAGTAAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1024
982 GAGAGAGTAAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1041
```

QY 1025 TGGGCGGACCGAAGCTGATGTAGAGGTTAGCAACCCCAAGCCCTGTATTGCAATTTGCC 1084
 DB 1042 TGGGCGGACCGAAGCTGATGTAGAGGTTAGCAACCCCAAGCCCTGTATTGCAATTTGCC 1101
 QY 1085 GCCAGCTCACAACAAGGTTTATGATTTGCTGTAGACCGCTGTGTAAGATGTGTTTCATG 1144
 DB 1102 GCCAGCTCACAACAAGGTTTATGATTTGCTGTAGACCGCTGTGTAAGATGTGTTTCATG 1161
 QY 1145 GCGATTGTGTGGGCAATTTCTAGAGCTGAGGAGGCTTTTGGAAAGGAATGGGGAAGCT 1204
 DB 1162 GCGATTGTGTGGGCAATTTCTAGAGCTGAGGAGGCTTTTGGAAAGGAATGGGGAAGCT 1221
 QY 1205 ATATCTGCGCCAACTGACCACTTCTGCAAGTGTAGAGATGATCTATTCAGAAACGGGAG 1264
 DB 1222 ATATCTGCGCCAACTGACCACTTCTGCAAGTGTAGAGATGATCTATTCAGAAACGGGAG 1281
 QY 1265 ATCAGCAGGAAGCTAAATGAGAGCTGTGAGATGTCTGATGAGCAACGATTTGTAACAATATAG 1324
 DB 1282 ATCAGCAGGAAGCTAAATGAGAGCTGTGAGATGTCTGATGAGCAACGATTTGTAACAATATAG 1341
 QY 1325 GAACAAATGAGACGAAGTCTAGCGAAGCCAAAGGATTTAAAGGTTAGATTTGAGAAAGCTG 1384
 DB 1342 GAACAAATGAGACGAAGTCTAGCGAAGCCAAAGGATTTAAAGGTTAGATTTGAGAAAGCTG 1401
 QY 1385 CAAATCCAAAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGTATAGAGGCGCTGTG 1444
 DB 1402 CAAATCCAAAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGTATAGAGGCGCTGTG 1461
 QY 1445 CCTCAAAATGATTTGGCCCGGGTGTCTGTCAAGTGGCGAGCCGACTCGGTGTATCTGCA 1504
 DB 1462 CCTCAAAATGATTTGGCCCGGGTGTCTGTCAAGTGGCGAGCCGACTCGGTGTATCTGCA 1521
 QY 1505 GTATGATCTGTATCTCTCAAAACGCGCGGAGCGAACAATGAATTTTAAAGCTCAAGTAAAG 1564
 DB 1522 GTATGATCTGTATCTCTCAAAACGCGCGGAGCGAACAATGAATTTTAAAGCTCAAGTAAAG 1581
 QY 1565 AACAGAGCCCAAGCCTTAAAGAAAGATGAAGTGAAGCCAGAGAACCCAGTCTTCCGA 1624
 DB 1582 AACAGAGCCCAAGCCTTAAAGAAAGATGAAGTGAAGCCAGAGAACCCAGTCTTCCGA 1641
 QY 1625 AATCGGCTGTCTGAGGCAAGTATTAATCTCTTCTGTGCAAGAGACAGCTTCCAGAA 1684
 DB 1642 AATCGGCTGTCTGAGGCAAGTATTAATCTCTTCTGTGCAAGAGACAGCTTCCAGAA 1701
 QY 1685 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCTGCGGAGGTGAACACTCCGGGA 1744
 DB 1702 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCTGCGGAGGTGAACACTCCGGGA 1761
 QY 1745 AGGAAGCAGCTTGTGAGAGCAGCAACGCTGTGAGGCGAGGATCAACAATTACATGAG 1804
 DB 1762 AGGAAGCAGCTTGTGAGAGCAGCAACGCTGTGAGGCGAGGATCAACAATTACATGAG 1821
 QY 1805 TAAAGCAGAAAGAACTGTCTCTCTCTGCGGCTGCACTGTGTATTAATGATATCAACC 1864
 DB 1822 TAAAGCAGAAAGAACTGTCTCTCTCTGCGGCTGCACTGTGTATTAATGATATCAACC 1881
 QY 1865 TAGGGGTGTGCTCTGAGACCCCTCCGCTCTTCTGATGATGATCAACCCCTGGGCGTGT 1924
 DB 1882 TAGGGGTGTGCTCTGAGACCCCTCCGCTCTTCTGATGATGATCAACCCCTGGGCGTGT 1941
 QY 1925 CAGGACTGAGAGTTGCAAGCTTGTGTTAAGCTGATCAAGACACCGGCTGACCATGAGC 1984
 DB 1942 CAGGACTGAGAGTTGCAAGCTTGTGTTAAGCTGATCAAGACACCGGCTGACCATGAGC 2001
 QY 1985 GGAAGACAGAGCCCATGTCCAGAGTCCCTCTGCTGCTGTGTTCATCCCTAGTCTGT 2044
 DB 2002 GGAAGACAGAGCCCATGTCCAGAGTCCCTCTGCTGCTGTGTTCATCCCTAGTCTGT 2061
 QY 2045 AGGACTTCCTGTCTCTGTTTCCAAAGCTGTAACTCACTGTGGAAGCTTCACTTAAT 2104
 DB 2062 AGGACTTCCTGTCTCTGTTTCCAAAGCTGTAACTCACTGTGGAAGCTTCACTTAAT 2121

QY 2105 GATTGATCTTAAATCTCTGTTTTCACCTGAGGCTCTGTAGATTTGTAATTCCTTC 2164
 DB 2122 GATTGATCTTAAATCTCTGTTTTCACCTGAGGCTCTGTAGATTTGTAATTCCTTC 2181
 QY 2165 ATCCAGTCTGATTCATAGCCACACTGCGCGGACGCGCAATCCACCCCTGTCTGACA 2224
 DB 2182 ATCCAGTCTGATTCATAGCCACACTGCGCGGACGCGCAATCCACCCCTGTCTGACA 2241
 QY 2225 TGAGTTGTTCTGACAAAGGCGGTATAGGCTGAGTTTTCACATTTGTCACAGGCGAG 2284
 DB 2242 TGAGTTGTTCTGACAAAGGCGGTATAGGCTGAGTTTTCACATTTGTCACAGGCGAG 2301
 QY 2285 CACATGAAGCATCACTCTTCTTATGTTGTTGAGGAACTTTGCAAGTATGTTGATC 2344
 DB 2302 CACATGAAGCATCACTCTTCTTATGTTGTTGAGGAACTTTGCAAGTATGTTGATC 2361
 QY 2345 TGATTTGAGGTGTACATTATTTTGTGCTGGGCAATAGGGGATTTTTTTTTTTCGA 2404
 DB 2362 TGATTTGAGGTGTACATTATTTTGTGCTGGGCAATAGGGGATTTTTTTTTTTCGA 2419
 QY 2405 TGTCCGATTTCACAGCTACACACCCACATGAACACATTCGAACTCGAAGG-CACACACT 2463
 DB 2420 TGTCCGATTTCACAGCTACACACCCACATGAACACATTCGAACTCGAAGGCAACACT 2479
 QY 2464 CCTGCTTCATAGGCGCCCAAGGTAAGTATCAACCTAGAACACTGTCTGTGACCGCAG 2523
 DB 2480 CCTGCTTCATAGGCGCCCAAGGTAAGTATCAACCTAGAACACTGTCTGTGACCGCAG 2539
 QY 2524 ACGGTGCTTTGAGCTTGTATTTCTACATGTGACTGAGCTTTTCTGCTCTCTTGA 2583
 DB 2540 ACGGTGCTTTGAGCTTGTATTTCTACATGTGACTGAGCTTTTCTGCTCTCTTGA 2599
 QY 2584 TGTTAGACTCTTAAGATCATATCTG 2610
 DB 2600 TGTTAGACTCTTAAGATCATATCTG 2626

RESULT 5

AX878224 2772 bp DNA linear PAT 17-DEC-2003
 LOCUS AX878224

DEFINITION Sequence 13129 from Patent EP1074617.
 ACCESSION AX878224

VERSION AX878224.1 GI:40032960
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE

1. Isegai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
 Iishi, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

TITLE Primers for synthesizing full-length cDNA and their use
 JOURNAL Patent: EP 1074617-A 13129 07-FEB-2001;

Research Association for Biotechnology (JP)
 FEATURES
 source

1. 2772
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

CDS
 301..1989
 /note="unassigned protein product"
 /codon_start=1
 /protein_id="CA890061.1"
 /db_xref="GI:40032961"

/translation="MDKSPNSNEARAKIKPTSKBRKXWGPRRTIARBGAGDAE
 ADLEPPPPQQLSLRSGRQPKRTERVEQFTIARRRRBRKPSLESGPTSC
 PATDARTASGVSASSTRSGPQASATAVARPARSSERKVGSDHDHDSGSLT
 LKELQNTLRREKEQPTERPLKGIOSLRKRREBGAFTVGSASDIVEGLSKOE
 PENQGVVSGKODRESKLEKKAQOIKEBEPDLRPEKREGYFNALYICROP
 HNNRPMTCRCRCEWTFGDCVIGISEARGLERNGEYICNCTILQVDETHSLAD
 QGEAKRPRGADGTCISGTIRIBKSSBDGIRKIRKRNPSKTKLKIROPYLEAP
 GASRCIGRCCHVAQPDVSYCSNDCLIAHAIATKFTLSGKEQPKRKRRKMPKPK
 SLPKCAQAGIKISVHKRPAPEKRTTVKAVVAVPARSEALGREAECSSSTPMSWD

QY 2165 ATCCAGTCTGATTCATAGCCACATGCGCCGACGCGCATCCACCCCTGTCTGCACA 2224
DB 2201 ATCCAGTCTGATTCATAGCCACATGCGCCGACGCGCATCCACCCCTGTCTGCACA 2260
QY 2225 TGAAGTGTCTGACACACGCGCTGTATACGTTTCAATTTTCCACATTTGCCAGCGCCAG 2284
DB 2261 TGAAGTGTCTGACACACGCGCTGTATACGTTTCCACATTTTCCAGCGCGCCAG 2320
QY 2285 CACATGAAGATCACTCTTTTATGTGTGGGAATCTTGAAGTATGTTGCATC 2344
DB 2321 CACATGAAGATCACTCTTTTATGTGTGGGAATCTTGAAGTATGTTGCATC 2380
QY 2345 TGAATTTCAAGGTATCACTTTTATGTTTGAAGTATGTTTTCAT 2404
DB 2381 TGAATTTCAAGGTATCACTTTTATGTTTGAAGTATGTTTTCAT 2438
QY 2405 TGTTCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAAG-CAACACT 2463
DB 2439 TGTTCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAAGCCACACT 2498
QY 2464 CCGCTTCATAGGCGCCACCGCTAGTGTGATTCACACCTGAAACATGCTCCGACCGCAG 2523
DB 2499 CCGCTTCATAGGCGCCACCGCTAGTGTGATTCACACCTGAAACATGCTCCGACCGCAG 2558
QY 2524 ACCGCTCCTTGAAGCTTGTATTTCTACATGATGATGCTTCTTGCCCTCGTCTTGA 2583
DB 2559 ACCGCTCCTTGAAGCTTGTATTTCTACATGATGATGCTTCTTGCCCTCGTCTTGA 2618
QY 2584 TGTTTAGACTTAAATGATATATCTTG 2610
DB 2619 TGTTTAGACTTAAATGATATATCTTG 2645

RESULT 6
BD157094 2772 bp DNA linear PAT 17-JAN-2003
LOCUS Primer for synthesizing full-length cDNA and use thereof.
DEFINITION BD157094
ACCESSION BD157094.1 GI:27863852
VERSION JP 2002191363-A/11937.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Bukaryaota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 2772)
AUTHORS Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Makamatsu,A., Nagai,K. and Otsuki,T.
TITL Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 11937 09-JUL-2002;
HELIIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/11937
PD 09-JUL-2002
PF 28-JUL-2000 JP 200028090
PI TOSHIO OCA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI MAKAMATSU, PI KUNICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC 10', C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key Location/Qualifiers (301). (1986).
FEATURES SOURCE 1..2772
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 96.4%; Score 2516.2; DB 6; Length 2772;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;
QY 65 GGAGCTTAACTCCAGGGGAACAGCCCTAGATATCTGATGTTGAAATACGAAGCCTG 124
DB 101 GGAGCTTAACTCCAGGGGAACAGCCCTAGATATCTGATGTTGAAATACGAAGCCTG 160
QY 125 TTAATCGGAACAGTGGCTGACAAAGTGTGTGTGAGCCCTGCTCTGGAACCC 184
DB 161 TTAATCGGAACAGTGGCTGACAAAGTGTGTGTGAGCCCTGCTCTGGAACCC 220
QY 185 AGAGGTTTCTGCTCCAGGGTTTGTGTGTATTTAGATTTCAAGGAAAGTGTCAAG 244
DB 221 AGAGGTTTCTGCTCCAGGGTTTGTGTGTATTTAGATTTCAAGGAAAGTGTCAAG 280
QY 245 CTTTCAGTGTGAGCAGATATGGAACAAGGCGAACCCGAGGAATAGAGAGGCACTTA 304
DB 281 CTTTCAGTGTGAGCAGATATGGAACAAGGCGAACCCGAGGAATAGAGAGGCACTTA 340
QY 305 AGGCATCAAAACCAACAGCAAAAGTTCAAGAAACATGGGGTTTTCGAAGACCACTA 364
DB 341 AGGCATCAAAACCAACAGCAAAAGTTCAAGAAACATGGGGTTTTCGAAGACCACTA 400
QY 365 TCGCAAGCGAGAGGCGCAGGGGAACGCGAGGCTGAATCCATGAGCCGCCACCCCAAC 424
DB 401 TCGCAAGCGAGAGGCGCAGGGGAACGCGAGGCTGAATCCATGAGCCGCCACCCCAAC 460
QY 425 AGCAGCAGCTGGGCTGTCCCTGGGGGCAATGGGAAGCAGCCCAAGGCGCACTGAGCGG 484
DB 461 AGCAGCAGCTGGGCTGTCCCTGGGGGCAATGGGAAGCAGCCCAAGGCGCACTGAGCGG 520
QY 485 TGAAGCAGTTCTGACCAATTCGCGCGCGCGCGCAGAGAGAGCATGCTGTCTCTG 544
DB 521 TGAAGCAGTTCTGACCAATTCGCGCGCGCGCGCAGAGAGAGCATGCTGTCTCTG 580
QY 545 AGGATTTGTGTGAGCCCACTGCTGCTCCCGCCACAGAGCCGACACCTCCAGAGGCA 604
DB 581 AGGATTTGTGTGAGCCCACTGCTGCTCCCGCCACAGAGCCGACACCTCCAGAGGCA 640
QY 605 GCGTGAAGAGGCTTCTGAGACCAAGAGGAGCCCGCAGTCTGCTTCCAGCTGTGAAG 664
DB 641 GCGTGAAGAGGCTTCTGAGACCAAGAGGAGCCCGCAGTCTGCTTCCAGCTGTGAAG 700
QY 665 AACGACAGCTCTTCTGAAAAGGTGAAGAGGAGTGAACACGATGACACTTCGATA 724
DB 701 AACGACAGCTCTTCTGAAAAGGTGAAGAGGAGTGAACACGATGACACTTCGATA 760
QY 725 GTGACAGGATGCTGACCTTGAAGAGCTTCAAGATGCTCTTGCAGAGAGCGGGAAC 784
DB 761 GTGACAGGATGCTGACCTTGAAGAGCTTCAAGATGCTCTTGCAGAGAGCGGGAAC 820
QY 785 AGGAGCCCACTGAGAGGCTTGAAGAGGATCCAGAGTGGCTGGGGAAGAGCGCGG 844
DB 821 AGGAGCCCACTGAGAGGCTTGAAGAGGATCCAGAGTGGCTGGGGAAGAGCGCGG 880
QY 845 AGGAGGCTCCCGCGAGACTGTGGGCTCCGAGGCCAGTGAACACTGTGAGAGGCGTCTG 904
DB 881 AGGAGGCTCCCGCGAGACTGTGGGCTCCGAGGCCAGTGAACACTGTGAGAGGCGTCTG 940
QY 905 CCAAGTAACAGAGGCCCGAGAACATCAAGGAGGTTGTCTCCAGGCTGGGAAAGTGA 964
DB 941 CCAAGTAACAGAGGCCCGAGAACATCAAGGAGGTTGTCTCCAGGCTGGGAAAGTGA 1000
QY 965 GAGAGATTAATTTGAGAGGAAAGCGGCTCAGAGCATGAAGATGAGAGGCTTGAGACT 1024
DB 1001 GAGAGATTAATTTGAGAGGAAAGCGGCTCAGAGCATGAAGATGAGAGGCTTGAGACT 1060
QY 1025 TGGGCGGACCGAAGCCTGAATGTGAGGTTACAGACCCCAAGCCCTGTATTGCAATTTGC 1084
DB 1061 TGGGCGGACCGAAGCCTGAATGTGAGGTTACAGACCCCAAGCCCTGTATTGCAATTTGC 1120
QY 1085 GCCAGCTTCACAAACAGGTTTATGATTGCTGTGACCGCTGTGAAAGATGTTCAATG 1144

Db 1121 GGCACCTCACAACAAGATTATGATTTGGCTGTACCGCTGTGAAGATGTTTCATG 1180
 Qy 1145 GCGATTGTGTGGGCAATTTCTGAGGCTCGAGGGGCTTTTGGAAAGAAATGGGAAAGACT 1204
 Db 1181 GCGATTGTGTGGGCAATTTCTGAGGCTCGAGGGGCTTTTGGAAAGAAATGGGAAAGACT 1240
 Qy 1205 AATATCGCCCAAACTGACCATTTCTGCAAGTGAAGATGAGCTCATTCAGAAACGGCAG 1264
 Db 1241 AATATCGCCCAAACTGACCATTTCTGCAAGTGAAGATGAGCTCATTCAGAAACGGCAG 1300
 Qy 1265 ATCAGCAGAGACTAATATGAGAGCTGAGAGATGATGAGCAACGATTTGATCAAGATATG 1324
 Db 1301 ATCAGCAGAGACTAATATGAGAGCTGAGAGATGATGAGCAACGATTTGATCAAGATATG 1360
 Qy 1335 GAAACAATAGAGAGAAAGTCTAGCGAAGACCAAGGATTAAGGGTAAATTTGAGAAAGCTG 1384
 Db 1361 GAAACAATAGAGAGAAAGTCTAGCGAAGACCAAGGATTAAGGGTAAATTTGAGAAAGCTG 1420
 Qy 1385 CAATTCAGAGTGGCAAGAGAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCTGTGTG 1444
 Db 1421 CAATTCAGAGTGGCAAGAGAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCTGTGTG 1480
 Qy 1445 CCTCAAAATGATATGTGCCCCGGGTGCTGTCACTGTGGCGAGCCCGA CTGGTGTACTGCA 1504
 Db 1481 CCTCAAAATGATATGTGCCCCGGGTGCTGTCACTGTGGCGAGCCCGA CTGGTGTACTGCA 1540
 Qy 1505 GTAATGACTGTATCTCTCAAAACGCGCGACGCAATGAAGTTCTTAAGCTCAGGTTAAG 1564
 Db 1541 GTAATGACTGTATCTCTCAAAACGCGCGACGCAATGAAGTTCTTAAGCTCAGGTTAAG 1600
 Qy 1565 AACAGAGCCAAAGGCTTAAGAAAGATGAAGTGAAGCAGAGAGCCCAAGCTTCCGA 1624
 Db 1601 AACAGAGCCAAAGGCTTAAGAAAGATGAAGTGAAGCAGAGAGCCCAAGCTTCCGA 1660
 Qy 1625 AATGCGGTGCTCAGGCGAGGTATTAATTAATCTCTTCTGTGCAACAAGAGCCAGCTCCAGAA 1684
 Db 1661 AATGCGGTGCTCAGGCGAGGTATTAATTAATCTCTTCTGTGCAACAAGAGCCAGCTCCAGAA 1720
 Qy 1685 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCTGCGGAGATGAAGCACTGGGA 1744
 Db 1721 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCTGCGGAGATGAAGCACTGGGA 1780
 Qy 1745 AGAAGACGCTGTGAGACGAGACGCGGTGCGGCGAGCGATCACAATTAACATGCAAG 1804
 Db 1781 AGAAGACGCTGTGAGACGAGACGCGGTGCGGCGAGCGATCACAATTAACATGCAAG 1840
 Qy 1805 TAAAGCCAGAAAAGACTGTGCTCCCTGCGCGGTCACTGTGTATTAATATGTATCAAC 1864
 Db 1841 TAAAGCCAGAAAAGACTGTGCTCCCTGCGCGGTCACTGTGTATTAATATGTATCAAC 1900
 Qy 1865 TAGGGGTGTGCTCTGTGACCCCTCCCTGCTTTCTGTGATAGCAATCCCTGTGGCTGTG 1924
 Db 1901 TAGGGGTGTGCTCTGTGACCCCTCCCTGCTTTCTGTGATAGCAATCCCTGTGGCTGTG 1960
 Qy 1925 CAGGACTGGAGATGTGAGCTTTGTGTAAGTGAATCAAGACACCGGCTGACATCAAG 1984
 Db 1961 CAGGACTGGAGATGTGAGCTTTGTGTAAGTGAATCAAGACACCGGCTGACATCAAG 2020
 Qy 1985 GGAAGACGAGAGCCATGTTCAGAGATGCTCTGTGCTGCTGTGTTCATCCCTAGTGTGTC 2044
 Db 2021 GGAAGACGAGAGCCATGTTCAGAGATGCTCTGTGCTGCTGTGTTCATCCCTAGTGTGTC 2080
 Qy 2045 AGGACTTCTGTGACTGTGTTTTCAGAAAGCTGTAACTCACTGTGAAAGCTTCACTTAAT 2104
 Db 2081 AGGACTTCTGTGACTGTGTTTTCAGAAAGCTGTAACTCACTGTGAAAGCTTCACTTAAT 2140
 Qy 2105 GATTGATCTTAATCTGTTTTCAGCTGCTGAGGCTGTGTAAGTATTTGATTTCTCTTC 2164
 Db 2141 GATTGATCTTAATCTGTTTTCAGCTGCTGAGGCTGTGTAAGTATTTGATTTCTCTTC 2200
 Qy 2165 ATCCAGTCTGATTCATAGCAACATGCGCGGACGCGACATCCACCTCTGTGCA 2224

Db 2201 ATCCAGTCTGATTCATAGCAACACTGCGCGGACGCGACATCCACCTCTGTGCA 2260
 Qy 2225 TAGGTTGTTCTGACACAGCGGCTGTATAGCTTCAGATTTTCCATGTCACGCGCAG 2284
 Db 2261 TAGGTTGTTCTGACACAGCGGCTGTATAGCTTCAGATTTTCCATGTCACGCGCAG 2320
 Qy 2285 CACATGAAGCATCACTCTTTTATATGTTGTGGAAATCTTTGCAAGTATGATGTGCATC 2344
 Db 2321 CACATGAAGCATCACTCTTTTATATGTTGTGGAAATCTTTGCAAGTATGATGTGCATC 2380
 Qy 2345 TGAATTTCAAGGTGTACATTTATTTTGTGACTGCGCAATGAGGGAATTTTTCCTCA 2404
 Db 2381 TGAATTTCAAGGTGTACATTTATTTTGTGACTGCGCAATGAGGGAATTTTTCCTCA 2438
 Qy 2405 TGTGCGATTCACAGCTTACACACCCACATGAAACATTTGAACTTTCGAAGG-CACACACT 2463
 Db 2439 TGTGCGATTCACAGCTTACACACCCACATGAAACATTTGAACTTTCGAAGGCGCACACT 2498
 Qy 2464 CTGCTTCATAGGCGCCACAGGTATGAGTTCACACTTGAACAACCTGTCTGACCGCAGG 2523
 Db 2499 CTGCTTCATAGGCGCCACAGGTATGAGTTCACACTTGAACAACCTGTCTGACCGCAGG 2558
 Qy 2524 AGCGTGCCTTGTGACTTGTGATTTCTACATGTGACTGTGCTTTCTTGCCCTGTCTGAA 2583
 Db 2559 AGCGTGCCTTGTGACTTGTGATTTCTACATGTGACTGTGCTTTCTTGCCCTGTCTGAA 2618
 Qy 2584 TGTGTAGACTCTTAAGATCATATCCCTG 2610
 Db 2619 TGTGTAGACTCTTAAGATCATATCCCTG 2645

RESULT 7
 AK002127
 LOCUS 2772 bp mRNA linear PRI 01-AUG-2002
 DEFINITION Homo sapiens CDNA FL111265 fis, clone PLACB1009158.
 ACCESSION AK002127
 VERSION AK002127.1 GI:7023814
 KEYWORDS oligo capping; file (full insert sequence).
 SOURCE
 ORGANISM Homo sapiens (human)

REFERENCE
 AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Magatsuna, M., Hosoi, T., Raku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chida, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuo, Y., Ninomiya, K. and Iwayanagi, T.

TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2772)
 AUTHORS Isogai, T. and Otsuki, T.
 TITLE Direct Submision
 JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES
 source
 Location/Qualifiers

1..2772
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACB1009158"
 /issue_type="Placenta"
 /clone_1ib="PLACB1"

CDS

/note="Cloning vector: pMB18F3"
301. 1989
/note="unamed protein product"
/codon_start=1
/protein_id="BAA92094.1"
/db_xref="GI:7023815"
/translation="MDDKSGPSNEAPKAIKPTSEPRKTKTFRRTTIAKREGADAE
ADLEPPPOQGLSLRSGROPKRTERVQFLTIARRGRSPVSLSDSEPTSC
PADLARTSSSVSASERTSGPOSASTAKERPASEKYGDHDDSDSDIGLT
LKEIQNLRKRKRQSPERPLKGIQSLRKRRREBPARTVGSASPTVGVLPKBE
PENDOGVSAQKQDRESKLEKKAODIKDEPEDLAPKPEKGYDPMALYCIKRP
HNHRFMICDRCEWFHGDICVIGISEARGLIERNGEDYICPNCITLQVDBTHETAD
QOEAKRMPGADTDCTSIGTIEOKSESDOIKRIBKAAPSKKDLKIFQPIEAP
GLKCKGPGCVAAQPDQSYCSNDCTIKMAATMKPISLSEKBOEKREKMKPERP
SLPCCAGACIKLISSEHRRPAPEKRTTVKAVVPPRSLGEBABCESTPMAD
HNVAVPERKTAAPSPSLTKCMTHLGVGLIDPSRSMIALPMACPLGVALAC"

ORIGIN

Query Match 96.4%; Score 2516.2; DB 9; Length 2772;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

65 GGAAGCTTACTCCAGGGAAACAGCTTATGATATCTGAGTTGTGAAATAAGAACCTG 124
101 GGAGCTTACTCCAGGGAAACAGCTTATGATATCTGAGTTGTGAAATAAGAACCTG 160
125 TTACTCGTGAACAGTGGCTGACAAAGTGTGTTGTGAGCTGCTGTCTGTTGAAGCC 184
161 TTACTCGTGAACAGTGGCTGACAAAGTGTGTTGTGAGCTGCTGTCTGTTGAAGCC 220
185 AGAGGTTTCGCTGACAGGCTTTTGTTGTTATTTAGATTCAAGGAAAGTGTCAAG 244
221 AGAGGTTTCGCTGACAGGCTTTTGTTGTTATTTAGATTCAAGGAAAGTGTCAAG 280
245 CTTTCAGTGTGAGCAGTATGACAAAGGCGACCCGACGATGAGAGGACCTTA 304
281 CTTTCAGTGTGAGCAGTATGACAAAGGCGACCCGACGATGAGAGGACCTTA 340
305 AGGCCATCAAAACCCACGACAAAGATTCAAGAAACATGGGGTTTTCGAAAGACACTA 364
341 AGGCCATCAAAACCCACGACAAAGATTCAAGAAACATGGGGTTTTCGAAAGACACTA 400
365 TCGCAGGCGAGAGGCGCAGGGAGCGGAGGCTGACCACTGAGGCGGCAACCCGAC 424
401 TCGCAGGCGAGAGGCGCAGGGAGCGGAGGCTGACCACTGAGGCGGCAACCCGAC 460
425 AGCAGCAGCTGGGCTGTCCCTGCGGCGCAGTGGGAGGACGCCAAGGCGACTGAGCGG 484
461 AGCAGCAGCTGGGCTGTCCCTGCGGCGCAGTGGGAGGACGCCAAGGCGACTGAGCGG 520
485 TGAAGCAGTTCCTGACCAATGGCGGCGCCCGCGGCGAGAGACATGCTCTTCCTGG 544
521 TGAAGCAGTTCCTGACCAATGGCGGCGCCCGCGGCGAGAGACATGCTCTTCCTGG 580
545 AGGATTCTGAGAGGCGCAAGTCGTCGCGCCCGCGACAGAGCGGAGCAGCTCCGAGGGCA 604
581 AGGATTCTGAGAGGCGCAAGTCGTCGCGCCCGCGACAGAGCGGAGCAGCTCCGAGGGCA 640
605 GCGTGAAGAGGCGCTTCTGAGACAGAGCGGCGCCCGCAGTCTTCCTTCACAGCTGTGAAG 664
641 GCGTGAAGAGGCGCTTCTGAGACAGAGCGGCGCCCGCAGTCTTCCTTCACAGCTGTGAAG 700
665 AACGACAGGCTCTTCTGAAAGGTGAAGAGAGGAGATGACCAATGACACCTTCGATTA 724
701 AACGACAGGCTCTTCTGAAAGGTGAAGAGGAGATGACCAATGACACCTTCGATTA 760
725 GTGACAGGAGTGGCTTCTGAGACAGAGCTTCAAGATCGCTTTCGACAGAGCGGAGAAC 784
761 GTGACAGGAGTGGCTTCTGAGACAGAGCTTCAAGATCGCTTTCGACAGAGCGGAGAAC 820
785 AGGAGCCCACTGAGAGGCGCCCTGAAAGGAGATCCAGAGTGCCTGCGGAGAGAGCGCGG 844
821 AGGAGCCCACTGAGAGGCGCCCTGAAAGGAGATCCAGAGTGCCTGCGGAGAGAGCGCGG 880

845 AGAGGGTCCGCGCAGAGACTGTGGGCTCCGAGGCCAGTGAACATGTGAGGCGCTCTGC 904
881 AGAGGGTCCGCGCAGAGACTGTGGGCTCCGAGGCCAGTGAACATGTGAGGCGCTCTGC 940
905 CCAATTAACAGAGAGCCCGAGAACATAGAGGGGTTGTCTCCAGGCTGGGAAAGATGACA 964
941 CCAATTAACAGAGAGCCCGAGAACATAGAGGGGTTGTCTCCAGGCTGGGAAAGATGACA 1000
965 GAGGAGTAAAGTTGAGAGGAAAGCGGCTCAGGACATCAAGATGAGAGGCTGAGACT 1024
1001 GAGGAGTAAAGTTGAGAGGAAAGCGGCTCAGGACATCAAGATGAGAGGCTGAGACT 1060
1025 TGGCCGACAGAGAGCTGAAATGTGAGGTTACAGACCCCAAGCCCTGATTTGCAATTTGCC 1084
1061 TGGCCGACAGAGAGCTGAAATGTGAGGTTACAGACCCCAAGCCCTGATTTGCAATTTGCC 1120
1085 GCGAGCTTCAACAAACAGGTTTATGATTTGCTGTGACCGCTGTGAAAGATGTTCATG 1144
1121 GCGAGCTTCAACAAACAGGTTTATGATTTGCTGTGACCGCTGTGAAAGATGTTCATG 1180
1145 GCGATTTGTGGGCAATTTCTGAGGCTGAGGAGGCTTTTGGAAAGGAAAGGAGTCA 1204
1181 GCGATTTGTGGGCAATTTCTGAGGCTGAGGAGGCTTTTGGAAAGGAAAGGAGTCA 1240
1205 ATATCTGCCCCAACTGACCAATCTGCAAGTGCAGATGAGACTCAATCAAGAAACGGGAG 1264
1241 ATATCTGCCCCAACTGACCAATCTGCAAGTGCAGATGAGACTCAATCAAGAAACGGGAG 1300
1265 ATCAGCAGAGAGCTAAATGAGAGACTGAGAGTGTGATGAGCAACGATTTGATCAATATAG 1324
1301 ATCAGCAGAGAGCTAAATGAGAGACTGAGAGTGTGATGAGCAACGATTTGATCAATATAG 1360
1325 GAAACATGAGACAGAGTCTTACGCAAGCCAAAGGATTAAGGTTGATGAGAAAGCTG 1384
1361 GAAACATGAGACAGAGTCTTACGCAAGCCAAAGGATTAAGGTTGATGAGAAAGCTG 1420
1385 CAAATCCAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGATGAGGCGCTGTGTG 1444
1421 CAAATCCAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGATGAGGCGCTGTGTG 1480
1445 CCTCAAAATGTATTTGCCCCGGGTGTGTGACGTGGCGCAGCCGACCTGCTGTATCTGCA 1504
1481 CCTCAAAATGTATTTGCCCCGGGTGTGTGACGTGGCGCAGCCGACCTGCTGTATCTGCA 1540
1505 GTATGACTGTATCTCTCAAAACAGCCCGGAGGCAATGAAGTTTCTTAAAGCTCAGGTTAAG 1564
1541 GTATGACTGTATCTCTCAAAACAGCCCGGAGGCAATGAAGTTTCTTAAAGCTCAGGTTAAG 1600
1565 AACGAAAGCCAAAGCTTAAAGAAAGATGAAGTGAAGCCAGAGAGCCAGTCTTCGGA 1624
1601 AACGAAAGCCAAAGCTTAAAGAAAGATGAAGTGAAGCCAGAGAGCCAGTCTTCGGA 1660
1625 AATGCGTGTCTCAGGCAAGTATTAATAATCTCTTCTGTGCAACAGAGCAAGCTTCAGAAA 1684
1661 AATGCGTGTCTCAGGCAAGTATTAATAATCTCTTCTGTGCAACAGAGCAAGCTTCAGAAA 1720
1685 AAAAAAGACCAAGTGAAGAGGAGGAGTGTGCTCCGCGGAGGTGAAGACTCAGGGA 1744
1721 AAAAAAGACCAAGTGAAGAGGAGGAGTGTGCTCCGCGGAGGTGAAGACTCAGGGA 1780
1745 AGGAGGAGGCTTTGAGAGAGCAAGCGGCTGTGGGAGAGGATCACAATTAATCAATGAC 1804
1781 AGGAGGAGGCTTTGAGAGAGCAAGCGGCTGTGGGAGAGGATCACAATTAATCAATGAC 1840
1805 TAAAGCCAGAAAAGACTGTGCTCCCTGCGCGTCACTGTGTATAATGATATGATCAAC 1864
1841 TAAAGCCAGAAAAGACTGTGCTCCCTGCGCGTCACTGTGTATAATGATATGATCAAC 1900
1865 TAGGGGTTGGCTCTGTGACCCCTTCGCGTCTTTCTGTGATAGCAATCCCTGGGCGCTGC 1924
1901 TAGGGGTTGGCTCTGTGACCCCTTCGCGTCTTTCTGTGATAGCAATCCCTGGGCGCTGC 1960

[illegible]

misc_feature

similar manner to that of the RING and FYVE domains"
/db_xref=CDD:pfam00528"

Query Match 74.4%; Score 1941; DB 9; Length 2407;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 2259; Conservative 0; Mismatches 0; Indels 288; Gaps 3;

65 GGAAGCTTACTCCAGGGGAAACAGCTCTAGATATCTGAGTTGTGAAATACGAAGCTTG 124
1 GGAAGCTTACTCCAGGGGAAACAGCTCTAGATATCTGAGTTGTGAAATACGAAGCTTG 60

125 TTACTCGGAAACAGTGGCTGACCAAGTGTGTGTGAGCTGGCTGCTGCTTGAGACC 184
61 TTACTCGGAAACAGTGGCTGACCAAGTGTGTGTGAGCTGGCTGCTGCTTGAGACC 120

185 AGAGGTTTCGTCTGCGCAGGGTTTGTGTTGATTTAGATTTCAGGAAAGTGTCCAG 244
121 AGAGGTTTCGTCTGCGCAGGGTTTGTGTTGATTTAGATTTCAGGAAAGTGTCCAG 180

245 CTTTCAGTGTGGAGCAGATATGACCAAGGAGCAACCCAGCAATGAGAGGCACTTA 304
181 CTTTCAGTGTGGAGCAGATATGACCAAGGAGCAACCCAGCAATGAGAGGCACTTA 240

305 AGGCATCAAAACCCAGCAGAAAGATTCAAGGAAACATGGGGTTTTCGAAAGACACTA 364
241 AGGCATCAAAACCCAGCAGAAAGATTCAAGGAAACATGGGGTTTTCGAAAGACACTA 300

365 TCGCCAAGCAGAGAGGCGCAGGGGACGCGAGAGCTGACCCACTGAGCCGCCACCCAC 424
301 TCGCCAAGCAGAGAGGCGCAGGGGACGCGAGAGCTGACCCACTGAGCCGCCACCCAC 360

425 AGCAGCAGCTGGGCTGTTCCTTCGCGGCGCATGTGGAGGCAAGCCCACTGAGCGCG 484
361 AGCAGCAGCTGGGCTGTTCCTTCGCGGCGCATGTGGAGGCAAGCCCACTGAGCGCG 420

485 TGGAGCAGTTCCTGACCAATTGGCGCGGCGCGCGAGGAGAGCATGCTGTCTCCCTG 544
421 TGGAGCAGTTCCTGACCAATTGGCGCGGCGCGCGAGGAGAGCATGCTGTCTCCCTG 480

545 AGGATTCGTGTGAGCCCAAGTCTGCGCCCGCCACAGACCGCCGAGACAGCTTCGAGGCA 604
481 AGGATTCGTGTGAGCCCAAGTCTGCGCCCGCCACAGACCGCCGAGACAGCTTCGAGGCA 540

605 GCGTGGAAAGGCGCTTCTGAGACCAAGAGCGCGCCCACTGCTTCCACAGCTGGAAAG 664
541 GCGTGGAAAGGCGCTTCTGAGACCAAGAGCGCGCCCACTGCTTCCACAGCTGGAAAG 600

665 AACGACCAAGCTTCTGAAAAAGGTGAAGAGAGGGATGACCAAGATGACACTCCGATTA 724
601 AACGACCAAGCTTCTGAAAAAGGTGAAGAGAGGGATGACCAAGATGACACTCCGATTA 660

725 GTGACAGCGATGAGCTGACCTTGAAGAGCTTCAAGATGCTTCCAGAGAAAGCGGAAAC 784
661 GTGACAGCGATGAGCTGACCTTGAAGAGCTTCAAGATGCTTCCAGAGAAAGCGGAAAC 720

785 AGGAGCCCACTGAGAGAGCGCTTGAAGAGGATTCAGAGTCCCTGCGGAGAAAGCGCCG 844
721 AGGAGCCCACTGAGAGAGCGCTTGAAGAGGATTCAGAGTCCCTGCGGAGAAAGCGCCG 780

845 AGGAGGCTCCGCGCAGAGCTGTGGGCTCCGAGGCGCAGTGAACACTGTGAGGGCGTCTG 904
781 AGGAGGCTCCGCGCAGAGCTGTGGGCTCCGAGGCGCAGTGAACACTGTGAGGGCGTCTG 840

905 CCAAGTAACAGAGAGCCCGAGAACGATCAGAGGAGTTGTCTCCAGGCTGAGAAAGATGACA 964
841 CCAAGTAACAGAGAGCCCGAGAACGATCAGAGGAGTTGTCTCCAGGCTGAGAAAGATGACA 900

965 GAGAGAGTAATTTGAGAGGAAAGCGGCTCAGAGACATCAAGATGAGAGAGCTGAGAGCT 1024
901 GAGAGAGTAATTTGAGAGGAAAGCGGCTCAGAGACATCAAGATGAGAGAGCTGAGAGCT 960

1025 TGGGCGGACCGAAGGCTGAATGTGAGGGTTACGACCCCAAGCGCTGTATTTGCAATTGCGC 1084

1025 TGGGCGGACCGAAGGCTGAATGTGAGGGTTACGACCCCAAGCGCTGTATTTGCAATTGCGC 1084
961 TGGGCGGACCGAAGGCTGAATGTGAGGGTTACGACCCCAAGCGCTGTATTTGCAATTGCGC 1020

1085 GCCAGCTTCACACACAGGTTTATGATTTTGTGTGAGACCGCTGTGTAAGAAATGCTTTCAG 1144
1021 GCCAGCTTCACACACAGGTTTATGATTTTGTGTGAGACCGCTGTGTAAGAAATGCTTTCAG 1080

1145 GCGATTGTGTGGGCACTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGAAATGGGAAAGACT 1204
1081 GCGATTGTGTGGGCACTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGAAATGGGAAAGACT 1140

1205 ATATTCGCCCAACTGACCAATTCGCAAGTGCAGAGATGAGACTCATTCAGAAACGGCAG 1264
1141 ATATTCGCCCAACTGACCAATTCGCAAGTGCAGAGATGAGACTCATTCAGAAACGGCAG 1200

1285 ATCAGCAGAAAGCTTAAATGAGACCTGAGAGATGCTGATGAGCAACGATTGTAACAAGTATG 1324
1201 ATCAGCAGAAAGCTTAAATGAGACCTGAGAGATGCTGATGAGCAACGATTGTAACAAGTATG 1260

1325 GAAACATAGAGCAGAAAGTCTAGCGAAGACCAAGGGATTAAGGGTGAATTGAGAAAGCTG 1384
1281 GAAACATAGAGCAGAAAGTCTAGCGAAGACCAAGGGATTAAGGGTGAATTGAGAAAGCTG 1320

1385 CAATTCAGATGGCAGAGAAAGAACTCAAGATCTTCAGCTGTGATAGAGCGCTGTG 1444
1321 CAATTCAGATGGCAGAGAAAGAACTCAAGATCTTCAGCTGTGATAGAGCGCTGTG 1380

1445 CCTCAAAATGTATTTGGCCCCGGGGTGTGTGACAGTGGCCGACGCCGACCTCGGTACTGCA 1504
1381 CCTCAAAATGTATTTGGCCCCGGGGTGTGTGACAGTGGCCGACGCCGACCTCGGTACTGCA 1440

1505 GTAATGACTGTATCTCTCAAAACAGCGCGCAGCAGATGAAGTTTCTAAGCTCAGGTAAG 1564
1441 GTAATGACTGTATCTCTCAAAACAGCGCGCAGCAGATGAAGTTTCTAAGCTCAGGTAAG 1500

1565 AACGAGACCAAGGCTTAAAGAAAGATGAAGATGAAGCCAGAGAAAGCCAGTCTTCCGA 1624
1501 AACGAGACCAAGGCTTAAAGAAAGATGAAGATGAAGCCAGAGAAAGCCAGTCTTCCGA 1560

1625 AATGCGGTCTCAGGCAAGATTAATAATCTTCTGTGTGCAAGAGACCAAGCTCCAGAA 1684
1561 AATGCGGTCTCAGGCAAGATTAATAATCTTCTGTGTGCAAGAGACCAAGCTCCAGAA 1620

1685 AAAAAGAGACCAAGTGAAGAAAGCAGTGTGTGCTTCCGCGAGTGAAGCACTTCGGGA 1744
1621 AAAAAGAGACCAAGTGAAGAAAGCAGTGTGTGCTTCCGCGAGTGAAGCACTTCGGGA 1680

1745 AGGAAGCAGCTTGTGAGACAGCAGCGCTGTGGGCGAGCATCAATTCATGCAAG 1804
1681 AGGAAGCAGCTTGTGAGACAGCAGCGCTGTGGGCGAGCATCAATTCATGCAAG 1740

1805 TAAAGCCAGAAAGACTGTCTGCTCCCTGCGGTCACTGTGTATTAATGATATGATCAAC 1864
1741 TAAAGCCAGAAAGACTGTCTGCTCCCTGCGGTCACTGTGTATTAAT----- 1788

1865 TAGGGGTTGACCTCTGAGACCCCTCCCGTTCTTGTGATAGCAATCCCTGCGCTGTG 1924
1789 ----- 1788

1925 CAGGACTGGAGTTGACGCTTGTGTAACTGATCAAGACACCGGCTGCAACATCAGC 1984
1789 ----- 1788

1985 GGAAGACAGAGCCCATGTGCAAGATGCTTCGTCGCTGTGTCATCCATGCTGTG 2044
1789 ----- 1788

2045 AGGACTTCGTGCACTGTTTTCAAAGCTGTAACTGCACTGTGAAAGTTCACCTTAAT 2104
1789 ----- 1788

2105 GATTGATTCCTTAATCTCTGTTTTCACCTCAGGCTCTGTGATGATTTGATTTCTTTC 2164
1789 ----- 1788

QY 740 TGAACCTTGAAGAGCTTCAGAAATGCGCTTGCGAGAGAGCGGAAACAGAGCCCACTGGA 799
 DB 698 TTACGTGAAAGAACTTCAGAAACGCGCTTCGAGAAAGCAGAGCAAAACCTTGAGGA 757
 QY 800 GGCCTCTGAAAGGATCCAGAGTCCCTTGCGAAAGAGCGCGGAGAGGGTCCCGCG 859
 DB 758 GGTCCCTGAGAGGAGTCAGAAATGCGCTGAGAGAAAGAGCGAGAGAAATCTGCG 817
 QY 860 AGACTGTGGCTCCGAGGCCAGTACACTGTGAGAGGCGCTCCGCACTAGAGAGC 919
 DB 818 AAATCGGAGGTCTCAAAATAGGCGAGTCCGAGCAGAGCAGACTCTCTGTATGAGAGAGC 877
 QY 920 CCGAGAAAGATCAAGAGGCTGTGTCAGAGCTGAGGAAAGATGACAGAGAGTAAGTGG 979
 DB 878 CTGAGGCTAGTCAGAGAGCAGATGTCCAGTCAGAGACAGATGACATGAAATCAGTTGG 937
 QY 980 AGGAAAGCGCGCTCAGAGCACTCAAGATGAGAGAGCTGAGAGCTGAGCGCAGCAGAGC 1039
 DB 938 AAGGAAAGGCGACTCAGAGGAAATACAGAGAAACCCAGGAAAGCGGCAAAACCAAGC 997
 QY 1040 CTGAATGTAGGGTTAGAGACCCCAAGCGCTGTATTTGATTTGCCCGCACTCAGACA 1099
 DB 998 CTGAGTGTAGAGTTAGAGACCCCAATGCGCTGTACTGTGCTGCGCAGCCTCAGACA 1057
 QY 1100 AAGGTTATGATTTGCTGTGAGCGCTGTGAGAGATGTTTATGAGGAGTTGTGTGGCA 1159
 DB 1058 ACGGTTATGATTTGCTGTGATGTGTGAGAGATGTTTCAATGTGTACTGTGTGGTA 1117
 QY 1160 TTTCTGAGAGCTCGAGAGAGGCTTTTGGAAAGAAATGGGAAAGACTATATCTGCCCAACT 1219
 DB 1118 TTTCTGAGAGCGCGAGGCGGCTCTCTGAGAAAGAAAGGAGAACTATCTGCCCAACT 1177
 QY 1220 GCACATTTCTGAGAGTCAAGATGAGAGTCACTTCAAGAAAGCGAGATGACAGAGAGTA 1279
 DB 1178 GCACATTTCTGAGAGTCAAGATGAGAGTCACTTCAAGAAAGCGAGATGAGAGAGTCTG 1237
 QY 1280 AATGAGAGCTCGAGAGTCTGATGAGAGCCGATTTGTAACAATATGAGAAACAATAGAGCA 1339
 DB 1238 GGTCAATCTGTGGGTGTGTATGAGAGCTGCAAGAGATGAGAGCACTGAGAGCA 1297
 QY 1340 AGTCTAGCGAGAGCAAGAGGATTAAGGATTAAGATTAAGAAAGCTGCAAAATCAGAGTCCA 1399
 DB 1298 AGTCCGAGAGAGCAAGAGGATTAAGGATTAAGATTAAGAGAGAGCAAGAGCA 1357
 QY 1400 AGAAGAAACTCAAGATCTTCCAGCTGTGATGAGAGCGCTGTGCTCAAAATGTATTTG 1459
 DB 1358 AGAAGAAACTCAAGATCTTCCAGCTGTGATGAGAGCGCTGTGCTCAAAATGTATTTG 1417
 QY 1460 GCGCCGGGTGTCTCAGGTGCGAGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1519
 DB 1418 GCGCTGGGTGTCTCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1477
 QY 1520 TCAAAACGCGCGAGAGCAATGAAAGTTTCTAGCTCAGGTAAAGAACAGAGCCAAAGC 1579
 DB 1478 TCAAAACGCGCGAGAGCAATGAAAGTTTCTAGCTCAGGTAAAGAACAGAGCCAAAGC 1537
 QY 1580 CTAAAGAAAGATGAAGATGAAGAGAGAGAGCCAGTCTCCGAAAGCGCTGCTGAGG 1639
 DB 1538 CCAAGAAAGAGTGAAGAGAGAGAGAGAGAGTTCAGTCTTCAAAAGCAGGTTCAGG 1597
 QY 1640 CAGGTATTAAGATCTCTTCTGTGCAAGAGAGAGAGCTCCAGAAAGAAAGAGAGAGCAG 1699
 DB 1598 TGGGATTAAGATCTCTTCTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
 QY 1700 TGAAGAGAGAGTGTGTCTCTGCGAGAGTGAAGCACTCGGAGAGAGAGAGAGAGAGAG 1759
 DB 1658 TGAAGAGAGAGTGTGTCTCTGCGAGAGTGAAGCACTCTGAGAGAGAGAGAGAGAGAG 1714
 QY 1760 AGAGCAG 1819
 DB 1715 AGAGCAG 1774
 QY 1820 CTG-----CTGCTCCCTGCGCTGCACTGTGTATTAATGTATGATCACTTGAAGAGG 1870

DB 1775 CAG 1834
 QY 1871 TTGGCTCTCTGAG 1930
 DB 1835 CTGGCTCTCTGAG 1889
 QY 1931 TGGAGTTGAG 1990
 DB 1890 CAG-----AGTCTGGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1940
 QY 1991 CAG 2050
 DB 1941 CAG 2000
 QY 2051 TCTGTCT 2088
 DB 2001 CTCTGGGTGTCT 2038

RESULT 10

BD244713

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

PC

Db 61 CAGTCCCACTGCTCCCTTGTGTTTGAAGCTCCGGAATCTTCTATTGAAGAACTGTG 120
 Qy 141 GCTGACAAACAGTGTGTGTGAGCCTGCTGTCTCTGACCCAGAGGTTTCTGTGCC 200
 Db 121 ACTGACAAACAGTGTGTGAG--GCTTGGCCGTCTGTGACC-----TGGGCC 166
 Qy 201 AGGGTTTTGTGTATTAGATT--CAGGAAAAAGTCCAGCTTTCACTGTGTGAG 259
 Db 167 CAGGTCTAATATTTTATTAGGATTTCCAGCCAAAGTTTCCAGCTTCACTGTGTGGA 226
 Qy 260 CAGGTATGAGAACAAAGGCAACCCAGCAATGAGAGAGCACTTAAGCCATCAAAACA 319
 Db 227 CAGGTATGAGAACAAAGGCACTGAGCAATGAGAGAGCAACCAAGCTATCAAAACA 286
 Qy 320 CCAGCAAAAGATTGAGAAACATGGGGTTTTGAGAGACCACTATCCCAAGCAGAGAG 379
 Db 287 CCAATTAGAGATTGAGAAACCTGGGGTTTTGAGAGACCACTATCCCAAGCAGAGAG 346
 Qy 380 GCGCAGGAGACGCGAGGCTGACCCACTGAGCCGCAACCCCAAGCAGAGAGCTGGGCC 439
 Db 347 GTGCAAGAGACACGAGGCGGACCCCACTGAGCAGCAAC-----ACAGCAGCATTAAC 400
 Qy 440 TGTCTCTGCGCGCAGTGTGAGAGCAGCCCAAGCCTGAGCGCTGTGAGAGAGTTCTTGA 499
 Db 401 TCTCTCTGCGCGCAGTGTGAGAGCAGCAACCAATCTGAGAGGTTAGAGAGTTCTTGA 460
 Qy 500 CCATTGGCGCGCGCGCGCAGAGAGAGCAATGCTGTCTCTCTGAGAGATTCTGTGAGAC 559
 Db 461 CCAGGTTTGGCGCGCGCGAGAGAGAGAAATGTCCGGTGTCTCTGAGAGATTCTGTGAGAC 520
 Qy 560 CCAGTCTCTGCGCGCGCAGAGAGAGCAAGCTCTCTGAGAGAGCGCTGTGAGAGAGCGCTT 619
 Db 521 CCAATCTTCCACAGTCACTGATGTGAGAGACAGCTTCCAGAGGAGCGCTGTGAGAGAGCT 580
 Qy 620 CTGAGACAGAGAGAGGCGCGCGCAGTGTCTCTCTGAGAGAGTGTGAGAGAGAGAGCTCTT 679
 Db 581 CTGAGATGAGAGAGGCGCGCTGTATCTGACTCTTGA---GGGAAACCAATCTCTGTCTCTT 637
 Qy 680 CTGAAAGGTTGAAAGAGAGAGAGTGAACAAGTGAACCTCCGATGTGTAAGAGAGAGAGC 739
 Db 638 CTGAAAGGTTGAAAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 697
 Qy 740 TGAACCTTGAAGAGCTTGAAGATCGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 799
 Db 698 TTACGTTTGAAGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
 Qy 800 GCGCCCTGAAAGAGATCGAGAGTGGCTGTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAG 859
 Db 758 GGTCTCTGAGAGAGAGTGAAGATCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817
 Qy 860 AGACTGTGGGCTCCGAGGCGAGTGAACAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 919
 Db 818 AATCTGAGAGAGTCTCAATTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 877
 Qy 920 CCGAGAAAGATCAGAGAGAGTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 979
 Db 878 CTGAGAGAGAGTGAAG 937
 Qy 980 AGGAAAG 1039
 Db 938 AAGGAAAG 997
 Qy 1040 CTGAAATGAGAGAGTGAAG 1099
 Db 998 CTGAGTGTGAGAGTGAAG 1057
 Qy 1100 ACAGGTTTATGATTTGTGTGAG 1159
 Db 1058 ACAGGTTTATGATTTGTGTGAG 1117
 Qy 1160 TTTCTGAGAGCTGAG 1219

Db 1118 TTTCTGAG 1177
 Qy 1220 GCACCATTTCTGAG 1279
 Db 1178 GCACCATTTCTGAG 1237
 Qy 1280 AATGAG 1339
 Db 1228 GGTGAG 1297
 Qy 1340 AGTCTGAG 1399
 Db 1298 AGTCTGAG 1357
 Qy 1400 AG 1459
 Db 1358 AG 1417
 Qy 1460 GCGCGCGCGCTGTGAG 1519
 Db 1418 GCGCTGGGGTGTCTGAG 1477
 Qy 1520 TCAACAG 1579
 Db 1478 TCAACAG 1537
 Qy 1580 CTAAAG 1639
 Db 1538 CCAAG 1597
 Qy 1640 CAGGATTTAAATCTCTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1699
 Db 1598 TGGGATTTAAATCTCTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
 Qy 1700 TGAAG 1759
 Db 1658 TGAAG 1714
 Qy 1760 AG 1819
 Db 1715 AG 1774
 Qy 1820 CTG-----CTGCTCTCTGCGCTGAGTGTGATTAATGATTAATGATTAATGATTAATGATTA 1870
 Db 1775 CAG 1834
 Qy 1871 TGGGCTCTGAG 1930
 Db 1835 CTGGCTCTGAG 1889
 Qy 1931 TGGGAGTTCAG 1990
 Db 1890 CAG-----AGTCTGGGGTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
 Qy 1991 CAG 2050
 Db 1941 CAG 2000
 Qy 2051 TCTCTGAG 2088
 Db 2001 CTCTGGGGTGTCTGAG 2038

RESULT 11

AX023370

LOCUS 2867 bp DNA linear PAT 15-SEP-2000

DEFINITION Sequence 2 from Patent WO0015787.

ACCESSION

AX023370

VERSION AX023370.1 GI:10183782

KEYWORDS

SOURCE

ORGANISM

Mus sp. Mus sp. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

REFERENCE 1
 AUTHORS Leonardo, B.; Martinez, P.; Alonso, C.M.; Domingo, D.G. and Grandien, A.
 TITLE Genes encoding for the human and murine death in
 ducer-obliterator-1
 Patent: WO 0015787-A 2 23-MAR-2000;
 JOURNAL LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR
 INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO
 CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALP (SB)
 FEATURES
 source 1..2867
 /organism="Mus sp. n."
 /mol_type="unassigned DNA"
 /db_xref="taxon:10095"
 ORIGIN
 Query Match 39.2%; Score 1023.6; DB 6; Length 2867;
 Best Local Similarity 71.9%; Pred. No. 8,7e-264;
 Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;
 QY 21 TCCGCGCGGTTCCGGGAAATGGCTGCGAGACCTTAGAGCGCTGCGAGCTTACTCCACGG 80
 DB 1 TCCGCGGAGCTCGGAAATGGCTGCGGATCCCGCGCGCGGAGCTTGTTAAAGG 60
 QY 81 GAACAGCTCTAGATATCTAGTTGTTGAATACGAAGCTCTTACTCTGTAACAGT 140
 DB 61 CAGTCCCACTGTCCTTGTGTGTTCGAGCTCCGGAATCTTCTATTGATGAATCTG 120
 QY 141 GCTGACACAGGTTGTTGTGAGCTGCTGCTGCTTGAACCCAGAGTTTCGCTGC 200
 DB 121 ACTACACAGATGGGGTAG--GCTTGGCCGTCTGCTTAC-----TGGCC 166
 QY 201 AGGGTTTTGTGTTATTAGGATT--CAGGAAAAGTTCACAGCTTCACTGTTGAG 259
 DB 167 CAGGCTATATTTATATGATGATTTCCAGCCAAAGTTTCCAGCTTCACTGTTGAG 226
 QY 260 CAGGTTTGAAGCAAAAGCCACCCGAGCAATGAGAGGCACTTAAGCCATCAACCA 319
 DB 227 CAGGTTTGAAGTAAAGGCGCTGAGCAATGAGAGGCAACCCAGGCTATCAACCA 286
 QY 320 CCAGCAAAAGATTGAGAAAACATGGGGTTTTCGAAAGACATATCCCAAGCAGAGG 379
 DB 287 CCAATGAAGATTGAGAAAACCTGGGTTTTCGAAAGACATATCCCAAGCAGAGG 346
 QY 380 GCGCAGGGAAGCGGAGGCTGACCCAGCTGAGCCGCAACCCCAAGAGAGCTGGGCC 439
 DB 347 GTGAGGAGACACGAGAGCGGAGCCCACTGAGCAGCAAC-----ACAGCAGCATAAC 400
 QY 440 TGTCTCTGCGCGCAGTGGAGGAGCAGCCCAAGCCGCACTGAGCGCTGAGAGCATTC 499
 DB 401 TCTCTCTGCGCGCAGTGGAGGAGCAGCAACCAACGTAAGAGGGGTAGAGAGTTCTTA 460
 QY 500 CCATTGCGCGGCGCGCGCAGAGAGAGCATGCTGTCTCCCTGAGGATTCTGTGAGC 559
 DB 461 CCAAGGTTCCGCGCGCGAGGAGAAAAGATGTCGCGTCTCTGAGAGATTCACTGAGC 520
 QY 560 CCAGCTCTGCGCGCGCAGACGACGACGCTCCGAGGCGAGCTGGAAGCGCTT 619
 DB 521 CCAATCTTCCAGATCACTGATGAGAGACGCTTCCAGGGGAGCTGTAACAGATT 580
 QY 620 CTGAGACCAAGAGCGGCCCCCAGTCTGCTTCCACGCTGTGAAGAGAACCAAGCTCT 679
 DB 581 CTGAGATCAGAGTGGCTCTGTATCTGACTCTTA--GGGAAAACATCTCTGCTCTT 637
 QY 680 CTGAAAAGGTGAAGAGGAGATGACACGATGACATCTCCGATATGATACAGCATGAGC 739
 DB 638 CTGAAAAGGCAAAAGAGAGTGAAGAGAGAGACATCTTGAAGTGAAGTGAAGGCG 697
 QY 740 TGACCTTGAAGAGCTTCAAGATGCTCTTCCGAGAGCGGGAACAGAGCTCACTGAG 799
 DB 698 TTACGTTGAAGAACTTCAAGAACGCTTCCGAGAGAGAGAGAGCAAGAGCTGAGAG 757
 QY 800 GGCCCTGAAGAGGATTCAGAGTCCCTGCGGAGAAAGCGCCGAGAGAGGCTCCGCG 859

DB 758 GGTCTCTAGAGGCACTGAGATCGCTCTGAGGAAAGAACGAGAGAGAGATTCGCGC 817
 QY 860 AGACTGTGGCTCCGAGGCGCAGTGCACCTGTGAGGCGCTCTCCCACTAAGACAGAC 919
 DB 818 AATCTGGAGTGTCCAAATAGGCAAGTGCAGGAGCAGACACTCTCTGTAGCAGAGC 877
 QY 920 CCGAAGACATCAGAGGGTGTGTCCCGAGGCTGGGAAGATGACAGAGAGATTAAGTTG 979
 DB 878 CTGAGGCTAGTCAGGAGACAGTGTCCAGTGAAGACAGATATCATAGAAAATCAGTTG 937
 QY 980 AGGGAAGCGGCTCAGGACATCAAAAGATGAGAGACCTGAGACTTGGCCGACGGAAC 1039
 DB 938 AAGGGAAGGCACTCAGGAAATATCAGAGAAAACCCAGGAGAACGGCAAAACCAAGC 997
 QY 1040 CTGATGTGAGGTTACGACCCCAACGCTGTATTCATTTGCGCCAGCTTCAACACA 1099
 DB 998 CTGAGTGTGAGGTTTACGACCCCAATGCTGTATCTGATCTGCGCCAGCTTCAACACA 1057
 QY 1100 ACAGGTTATATTTGCTGTGACCGCTGTGAAGATGGTTTCAATGCGCATGTGTGGGCA 1159
 DB 1058 ACAGGTTATATCTGCTGTATGCTGTGAGAGTGTTCATGCTGACTGTGGGTA 1117
 QY 1160 TTTGAGGCTCGAGGAGGCTTTTGAAGAGATGGGGAAGACTATATCTGCCCAACT 1219
 DB 1118 TTTCTGAGGCTCGAGGAGGCTCTCTGAAAGAGACGGGGAAGACTATATCTGCCCAACT 1177
 QY 1220 GCACCATTTCTGCACTGAGAGATGAGACTATTCAAGAAACGCGACATCAGCAGAACTA 1279
 DB 1178 GCACCATTTTGCAGATGCAAGATGAGACAAACGATAGCGCCACCAATGAGCAGACTG 1237
 QY 1280 AATGAGACTGTGAGATCTGTATGAGCAACCAATGATCAAGATATAGAAACATATGAGCAGA 1339
 DB 1238 GGTGCAATCTGTGTGTCTGTATGAGCAAGACTGCAAGACATAGGAGACATGAGACAGA 1297
 QY 1340 AGTTCAGGAGAGACCAAGGATTAAGGTTAGAAATGAGAAAGCTCAATTCAGTGGCA 1399
 DB 1298 AGTCCGAGAGAGACCAAGGATTAAGGTTAGAAATGAGACCAAAACCCAGGCGCA 1357
 QY 1400 AGAAGAACTCAAGATCTTCCAGCTGTGATGAGCGCGCTGTGCTTCAAAATGATTTG 1459
 DB 1358 AGAAAACCTCAAGATTTCCAGCTGTGTGATGAGCGCTGTGCTTCAAAATGATTTG 1417
 QY 1460 GCCCGGGTGTCTGACCTGTGGCGGAGCCGCACTGCTGTATGAGATGATGATTC 1519
 DB 1418 GCCCTGGGTTTCAAGTGTAGCAGCTGATCTGTGTATGAGATGATGATTC 1477
 QY 1520 TCAACACGCGCAGCAGCAATGAGATTCTAAGCTCAAGTTAAAGACAGAAC 1579
 DB 1478 TCAACACGCGCAGCAGCTACATGAGATTTCTAAGTTCAAGTTAAAGACAAAAAC 1537
 QY 1580 CTAAAGAAAGATGAGATGAGACCAAGAACCCAGTCTTCCGAATGCGGTCTCAG 1639
 DB 1538 CCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAAGTCTTCCAAAATGCACTGTTCAG 1597
 QY 1640 CAGGATTAATAATCTCTCTGTGCAAGAGAGCCAGCTCCAGAAAAGAGAGACCAAG 1699
 DB 1598 TGGGATTAATAATCTCTCTGTGCAAGAGAGCTAGCTCAAGAAAAGAGAGAACCCAG 1657
 QY 1700 TGAAGAGGAGTGTGCTCCCTGCGCGAGTGAAGCACTCGGAGAGGAAGCAGCTTGTG 1759
 DB 1658 TGAAGAA---AGTATGCTGCGCTTCAGAGAGTGAACATTTGGAAGAGGAGCAGCTGTG 1714
 QY 1760 AGAGCAGCAACGCTGTGCGGCGAGCGCATCAATATTAACATGCAATTAAGCCAGAAAAG 1819
 DB 1715 AGAGCAGCAACATCTGCGGCGAGAGTGAACCAATTAACATGCAATTAAGCCAGAAAAG 1774
 QY 1820 CTG-----CTGCTCTGCGCGGCACTGTATTAATATGATATCACTTACGAGG 1870
 DB 1775 CAGAGAGCCCACTGCACTCTGCGCCACCTATTAAGTAATATGATATCACTTACGAGG 1834
 QY 1871 TTGGCTCTGAGACCTCTCCGTTCTTTCTGATGAGCAATCCCTGGGCTGTCAAGAC 1930

Db 1835 CTGGCTTCCAGGCCCTCCATCATCT-...-GGTGGCTGCTGGGGCTGTCTAGAC 1889

Qy 1931 TGGAGTTCAGAGCTTTGTGTTAAGCTATCAGACACCGGCTGCACCATCAGCGGAG 1990

Db 1890 CAG-----AGTCTGGGTGTTCTGTCTGTATAGTACGACGACC-----TACGTCCAGC 1940

Qy 1991 CAGAGCCATGTCCAGAGATGCTCTGCTGCTGCTGTGTCCATCCCTTACTGTCTAGACT 2050

Db 1941 CAGAGCAGATACCAAGATGCTCTGAGACCCAGGTGTTCTGCTTACGCTGTGAGCCT 2000

Qy 2051 TCGTGTACCTGTTTCCAAAGCTGTAACCTCAGCTGT 2088

Db 2001 CTCTGGGTGTTCTCTAAAGAGCTGTAGGCTCATGT 2038

RESULT 12

AKI29117 4906 bp mRNA linear ROD 21-NOV-2003

LOCUS DEFINITION Mus musculus mRNA for mKIAA0333 protein.

AKI29117

AKI29117.1 GI:37359897

VERSION FLI CDNA.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hirooka, S., Suga, Y., Nagase, T., Ohara, O. and Koga, H.

TITLE Prediction of the coding sequences of mouse homologues of KIAA gene: III. the complete nucleotide sequences of 500 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries

JOURNAL DNA Res. 10 (4), 167-180 (2003)

MEDLINE 22972043

PUBMED 14621295

2 (bases 1 to 4906)

Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.

REFERENCE Direct Submission

AUTHORS Submitted (23-JUL-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)

TITLE The CREAT program supported by Japan science and technology corporation; cDNA full insert sequencing; Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.

COMMENT 3'-end one pass sequencing.

FEATURES

Source

1..4906

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="mbg16015"

/tissue_type="brain"

/dev_stage="adult"

/note="vector:modified pBC SK+"

1..4906

/gene="mKIAA0333"

<224.3829

/gene="mKIAA0333"

/note="CDS is predicted by in silico analysis. Start codon is not identified."

/evidence=not experimental

/codon_start=1

/product="mKIAA0333 protein"

/protein_id="BAC37927.1"

/db_xref="GI:37359898"

/translation="FYVGFPAKQAFSGTGMDDKHLNSNEAPKAIKPTSEKPKRTWGRRTTIAREGAGDTEVPSBOQPOOHMISLRSGRQPKRTREVEPLTVRRKKNVPSLESDSEPTSTVTVETVETBSGVSSTIRSGPVSLSGKEHPSSEKAGKEREEDTSDSDGLTLEKLNRLARRRGRVRSLSGSLRLEKRRERASATGVSQTSABDRFLCCKRCPASGPPVSOSTDIDENLEKATGATNTENPREKRPKPEC EYDPPNALYCIQRPHNNRMICCDRCBEMFHDGVCISERGLLRNGDYICPNC

gene

CDS

ORIGIN

Query Match 37.7%; Score 985; DB 10; Length 4906;

Best Local Similarity 76.0%; Pred. No. 2,3e-253; Indels 22; Gaps 5;

Matches 1288; Conservative 0; Mismatches 385;

Qy 175 GCTTGAACCAAGAGTTGCTGCTGAGGCTTTTGTGTATTTAGATTT-CAGGAA 233

Db 187 GCTTGGCGTGTGCTGACCTGAGCTTATATTTATATTTATGATTTCCAGCCAA 246

Qy 234 AAGTTCAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 293

Db 247 AGTTTCCAGCTTACAGTGTGAGCAGTATGATATTAAGGAGCCTGAGCATTA 306

Qy 294 GAGGACCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 353

Db 307 GAGGACCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 366

Qy 354 AAGGACCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 413

Db 367 AAGGACCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 426

Qy 414 GCCACCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 473

Db 427 GCAACCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 480

Qy 474 CAGTACCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 533

Db 481 TACTGAGGAGTATGAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 540

Qy 534 TATCTCCCTGAGAGTGTGAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 593

Db 541 GATGCTTCCGAGAGTGTGAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 600

Qy 594 CTCCAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 653

Db 601 TTCCAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 660

Qy 654 AGCTGTGAGAGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 713

Db 661 A--GGAAAGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 717

Qy 714 CACCTCCATGATGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 773

Db 718 CACCTCCATGATGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 777

Qy 774 GAGGAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 833

Db 778 AAGGAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 837

Qy 834 GAGGAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 893

Db 838 GAGGAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 897

Qy 894 GAGGAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 953

Db 898 GAGGAGGAGCAGCTTACAGCTTACAGTGTGAGCAGTATGAGCACAAGAGCCAGCCAGCATTA 957

```

QY 954 GAAAGATGACGAGAGAGTAAAGTTGGAGGAGAAAGCCGCTCAGACATCAAGATGAGGA 1013
DB 958 GACAGATGACATGAGAAATCAGTTGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1017
QY 1014 GCTTGAGACTTGGGGCCGAGCCGAGAGCTGATGAGGGGTACACACCCCAAGCCCTGTATA 1073
DB 1018 CCCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1077
QY 1074 TTGCATTTGCGCGCAGCTCAGCAACAAGCTTATGATTTGCTGTGACCGCTGTGAGA 1133
DB 1078 CTGCATCTGCGCGCAGCTCAGCAACAAGCTTATGATTTGCTGTGACCGCTGTGAGA 1137
QY 1134 ATGTTTATGAGCGATGTGTGGGCAATTTCTGAGAGCTCGAGGAGGCTTTTGGAGAGGA 1193
DB 1138 GTGTTTATGAGCGATGTGTGGGCAATTTCTGAGAGCTCGAGGAGGCTTTTGGAGAGGA 1197
QY 1194 TGGGAGAGACTATATCTGAGCCAAATGCAACCATTTCTGCAAGTGTGAGATGAGACTATTC 1253
DB 1198 CCGGAGAGACTATATCTGAGCCAAATTTGCAACCATTTTGCAGATGTGAGATGAGAGAGCG 1257
QY 1254 AGAAACGAGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1313
DB 1258 TAGCGCAGCAGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1317
QY 1314 TACAGATTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1373
DB 1318 CACAGATTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1377
QY 1374 TGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1433
DB 1378 TGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1437
QY 1434 GGGCGCTGTGCTCAAAATGATTTGGCGCGGTGCTGACGTTGGGCGAGCCGAGCTC 1493
DB 1438 GGGCGCTGTGCTCAAAATGATTTGGCGCGGTGCTGACGTTGGGCGAGCCGAGCTC 1497
QY 1494 GGTCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1553
DB 1498 TGTGTATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1557
QY 1554 CTCAGTAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1613
DB 1558 TTCAGTAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1617
QY 1614 CAGTCTTCCGAATGCGGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1673
DB 1618 CAGTCTTCCGAATGCGGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1677
QY 1674 AGCTCCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1733
DB 1678 AGCTCCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1734
QY 1734 AGCACTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1793
DB 1738 GACTCTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1794
QY 1794 TTACAGATGAGTAAAGCCAGAAAGACTG-----CTGCTCCCTGCGCGTCACTGTT 1844
DB 1798 CTACAGATGAGTAAAGCCAGAAAGACTG-----CTGCTCCCTGCGCGTCACTGTT 1844
QY 1845 GTATTAATGTATGTA 1859
DB 1855 GAGTAAATCCATGAA 1869

```

```

RESULT 13
LOCUS HS885L7/c 160241 bp DNA linear PRI 19-APR-2001
DEFINITION Human DNA sequence from clone RPS-885L7 on chromosome 20q13.2-13.33
Contains ESTs, STSs, GSSs and eight CpG islands. Contains the 3'
end of the NTR1 gene for high affinity neurotensin receptor 1, a
putative novel gene, a novel gene similar to a fly gene, the gene
for opiod growth factor receptor (7-60 protein), the COL9A3 gene

```

for collagen IX alpha 3, a putative novel gene similar to a fly gene, the TCPL5 gene for basic helix-loop-helix transcription factor-like 5, an ARF4 (ADP-ribosylation factor 4) pseudogene, a novel gene and the 3' end of the DATF1 gene encoding death associated transcription factor 1, complete sequence.

AL035669
 AL035669.43 GI:8979786
 HNG: 7-60; ARF4; COL9A3; collagen; CpG island; DIO-1; KIAA0333; neurotensin; NTR1; opiod growth factor receptor; TCPL5.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 160241)
 Smith.M.
 Direct Submission
 Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jul 8, 2000 this sequence version replaced gi:8919619.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Mp, MORNREP; Information on the MORNREP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormrep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>
 This sequence is the entire insert of clone RPS-885L7. The true left end of clone RPS-885L7 is at 140192 in this sequence. The right end of clone RPS-885L7 is at 44512 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPS-885L7 is from the library RPS1-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2.

FEATURES

source

1. 160241
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="q13.2-13.33"
 /clone="RPS-885L7"
 /clone_id="RPS1-5"
 2. 64

repeat_region

repeat_region

repeat_region

repeat_region

gene

mRNA

65..353
 /note="Alu repeat: matches 1. 287 of consensus"
 354..397
 /note="L1M9 repeat: matches 6254. 6302 of consensus"
 3862..11948
 /gene="NTR1"
 join(<3862..4063,7443..7533,9195..11948)
 /gene="NTR1"
 /product="NTR1 protein"
 /note="NTR1 protein (neurotensin receptor 1 (high affinity)), isoform 1"
 /note="match: CDNA: Em:AB017027 Em:E11480 Em:X70070 match: ESTs: Em:AA078742 Em:U47726 Em:H29546 Em:AA157908 Em:AA074404 Em:H29638 Em:AA079834 Em:AA157852"

```

CDS
/evidence=not_experimental
join(3862..4063,7443..7533,9195..9444)
/gene="NTSR1"
/note="Continues in BA93B14 (Em:ALJ57033) as BA93B14.3
match: proteins: Tr:O88319 Sw:P30989 Sw:P20789"
/codon_start=1
/evidence=not_experimental
/product="d885L7.1.1 (neurotensin receptor 1 (high
affinity), isoform 1)"
/protein_id="CAC12747.1"
/db_xref="GI:12313997"
/db_xref="GOA:P30989"
/db_xref="SWISS-PROT:P30989"
/translation="VNTFMSRTFPMNVIVSVLTATITANKLTVMRQAAEQGYCTGGE
HSTFSMAJEPGVQALRHGVRLAVVIAFVVCMLPVHRLMPCYISDQWTPFLYD
FYHYFYMTALFYVSTINPLTINLVANFRIPLATLACLPVRRRRRRKPAFSRK
ADSVSNHTLSNATRETTY"
repeat_region
4198..5157
/note="12 copies 80 mer 61% conserved"
misc_feature
complement(4223..5153)
/note="Random repeat. region contains two forced joins and
single clone. Assembly consistent with restriction
digest."
repeat_region
4223..5114
/note="443 copies 2 mer cc 59% conserved"
repeat_region
4231..5166
/note="234 copies 4 mer tccc 60% conserved"
repeat_region
4240..5151
/note="19 copies 48 mer 61% conserved"
repeat_region
4248..5159
/note="38 copies 24 mer 61% conserved"
repeat_region
4264..5163
/note="15 copies 60 mer 61% conserved"
repeat_region
4286..5152
/note="17 copies 51 mer 59% conserved"
repeat_region
4308..5171
/note="27 copies 32 mer 61% conserved"
repeat_region
4317..5156
/note="21 copies 40 mer 62% conserved"
repeat_region
4344..5142
/note="17 copies 47 mer 59% conserved"
repeat_region
4348..5145
/note="42 copies 19 mer 60% conserved"
repeat_region
4431..5144
/note="21 copies 34 mer 60% conserved"
repeat_region
4456..5148
/note="21 copies 33 mer 60% conserved"
repeat_region
4486..5151
/note="9 copies 74 mer 69% conserved"
repeat_region
4530..5087
/note="18 copies 31 mer 71% conserved"
repeat_region
4533..4822
/note="5 copies 58 mer 77% conserved"
repeat_region
4720..5137
/note="11 copies 38 mer 61% conserved"
repeat_region
4833..5064
/note="4 copies 58 mer 79% conserved"
repeat_region
5051..5164
/note="19 copies 6 mer tcttc 63% conserved"
join(5808..5912,7443..7533,9195..11948)
/gene="NTSR1"
/product="d885L7.1.2 (neurotensin receptor 1 (high
affinity), isoform 2)"
/note="match: ESTs: Em:ALJ57553"
/evidence=not_experimental
5934..6864
/note="LIM4 repeat: matches 4509..5488 of consensus"
repeat_region
6974..7049
/note="4 copies 19 mer 80% conserved"
repeat_region
7224..7451
/note="6 copies 38 mer 98% conserved"
repeat_region
7892..8074
/note="MER68A repeat: matches 68..210 of consensus"

```

```

repeat_region
8674..8966
/note="AluY repeat: matches 5..296 of consensus"
repeat_region
10876..10962
/note="MIR repeat: matches 86..162 of consensus"
polyA_signal
11918..11923
/gene="NTSR1"
11948
polyA_site
11948
/gene="NTSR1"
11974..12025
/note="13 copies 4 mer ctca 73% conserved"
repeat_region
14394..14478
/note="5 copies 17 mer 77% conserved"
misc_feature
14766..14966
/note="match: GSS: Em:B99679"
repeat_region
15824..15891
/note="17 copies 4 mer tccc 67% conserved"
repeat_region
16266..16577
/note="AluSc repeat: matches 1..313 of consensus"
repeat_region
16578..16866
/note="17 copies 17 mer 65% conserved"
repeat_region
16581..16844
/note="11 copies 24 mer 68% conserved"
repeat_region
16582..16891
/note="10 copies 31 mer 66% conserved"
repeat_region
16583..16879
/note="9 copies 33 mer 66% conserved"
repeat_region
16585..16840
/note="16 copies 16 mer 66% conserved"
repeat_region
16586..16875
/note="5 copies 58 mer 72% conserved"
repeat_region
16590..16874
/note="15 copies 19 mer 64% conserved"
repeat_region
16593..16872
/note="7 copies 40 mer 67% conserved"
repeat_region
16598..16837
/note="5 copies 48 mer 73% conserved"
repeat_region
16603..16842
/note="3 copies 80 mer 77% conserved"
repeat_region
16638..16901
/note="44 copies 6 mer aaagaa 62% conserved"
repeat_region
16664..16885
/note="5 copies 74 mer 78% conserved"
repeat_region
16670..16904
/note="5 copies 47 mer 77% conserved"
repeat_region
16848..16839
/note="13 copies 4 mer gaag 100% conserved"
repeat_region
17287..17411
/note="MER5A repeat: matches 35..165 of consensus"

Query Match      36.3%; Score 948.4; DB 9; Length 160241;
Best Local Similarity 99.6%; Pred. No. 36-243;
Matches 972; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 1636 CAGGACGATTAATCTCTTGACAGAGACGCTCCAGAAAAAGAGACC 1695
DB 155280 CAGGACGATTAATCTCTTGACAGAGACGCTCCAGAAAAAGAGACC 155221
QY 1696 ACAGTGAAGAGACGATGTCCTCGCGGAGTGAACACTCGGAGAGAGACGCT 1755
DB 155220 ACAGTGAAGAGACGATGTCCTCGCGGAGTGAACACTCGGAGAGAGACGCT 155161
QY 1756 TGTGAGAGACGACGCGCTGTCGGGAGACGATCAATTAATGATCACTAAGCCAGAA 1815
DB 155160 TGTGAGAGACGACGCGCTGTCGGGAGACGATCAATTAATGATCACTAAGCCAGAA 155101
QY 1816 AAGACTGCTGCTCCGCGCTGCTGATTAATGATGATCACTAAGGGTGGC 1875
DB 155100 AAGACTGCTGCTCCGCGCTGCTGATTAATGATGATCACTAAGGGTGGC 155041
QY 1876 CTCCTGACCCCTCCGCTCTTCTGATAGACCATCCCTGAGCCTGTCAGACTGGGA 1935
DB 155040 CTCCTGACCCCTCCGCTCTTCTGATAGACCATCCCTGAGCCTGTCAGACTGGGA 154981

```



```

Oy 1426 ----- 1425
Db 481 CCAGCTCCTGCTCTCTGAGGATGGAATTGCTGTGTCTGAAGACATCTCTGCT 540
Oy 1426 ----- GTGATAGAGGCGCTGCTGCTCAAAATGTA 1456
Db 541 TCACACTCTTGACATGACATTAGCTGACAGGTATAGAGCGCTGCTGCTCAAAATGTA 600
Oy 1457 TTGGCCCCGGGTGCTGTCACTGTGGGAGCCGCACTCGGTGACTGACATTAATGTA 1516
Db 601 TTGGCCCCGGGTGCTGTCACTGTGGGAGCCGCACTCGGTGACTGACATTAATGTA 660
Oy 1517 TCTTCAAAACAGCCGACGACATGAACTTCTTAAGCTCAGGTAAAGAACAGACCA 1576
Db 661 TCTTCAAAACAGCCGACGACATGAACTTCTTAAGCTCAGGTAAAGAACAGACCA 720
Oy 1577 AGCCTAAAGAAAGTAAAGAAAGCAAGACCCGAGTCTTCCGAAATGCGGTGCTC 1636
Db 721 AGCCTAAAGAAAGTAAAGAAAGCAAGACCCGAGTCTTCCGAAATGCGGTGCTC 780
Oy 1637 AGGCAAGATTAAATCTCTTCTGTGCAACAGACACAGCTCCAGAAAAAAGAGACCA 1696
Db 781 AGGCAAGATTAAATCTCTTCTGTGCAACAGACACAGCTCCAGAAAAAAGAGACCA 840
Oy 1697 CAGTAAAGAAAGGAGTGTGTCTCTGCGGAGTGAAGCACTCGGAAAGAAAGCACTT 1756
Db 841 CAGTAAAGAAAGGAGTGTGTCTCTGCGGAGTGAAGCACTCGGAAAGAAAGCACTT 900
Oy 1757 GTGAAGACAGACCGCGCTGTGTGGGAGGATCAAAATTAAGCAATGAAGCCAGAA 1816
Db 901 GTGAAGACAGACCGCGCTGTGTGGGAGGATCAAAATTAAGCAATGAAGCCAGAA 960
Oy 1817 AGACTGCTGCTCTCTGCGCTCACTTTTATTAATCCACGAA 1859
Db 961 AGACTGCTGCTCTCTGCGCTCACTTTTATTAATCCACGAA 1003

```

```

RESULT 15
LOCUS HSJ563B14/28519 bp DNA linear PRI 19-APR-2001
DEFINITION Human DNA sequence from clone RP4-563B14 on chromosome 20. Contains the 5' of the DATP1 gene encoding the death associated transcription factor 1, the 5' end of a novel gene, ESTs, STSs, GSSs and four Cpg islands, complete sequence.
ACCESSION AL117379
VERSION AL117379.14 GI:7329921
KEYWORDS HTG; Cpg island; DATP1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 28519)
REFERENCE 1
AUTHORS Collier, R.
TITLE Direct Submissiion
JOURNAL Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1BA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Mar 26, 2000 this sequence version replaced gi:7329921.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submissiion corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPP; Information on the WORMPP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpp
This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at

```

FEATURES

```

source
    location/Qualifiers
    1..28519
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="20"
    /clone_lib="RP4-563B14"
    /clone_1bp="RP4-4"
    complement(1..26931)
    /gene="DATP1"
    /product="d5563B14.1 (Death associated transcription factor 1 (contains KIAA0333) isoform 1)"
    /note="match: cDNAs: Em:AK002127 Em:AJ238332 match: ESTs: Em:AA010673 Em:AA050179 Em:AA43704 Em:N55493"
    /evidence=not experimental
    complement(1..648)
    /gene="DATP1"
    /note="continues in clone d5885L7 (Em:AL035669) as gene d5885L7.9"
    match: proteins: Tr:Q9WV00"
    /codon_start=1
    /evidence=not experimental
    /product="d5563B14.1 (Death associated transcription factor 1 (contains KIAA0333) isoform 1)"
    /protein_id="CA90275.1"
    /db_xref="GI:7711029"
    /db_xref="GOA:Q9BTC0"
    /db_xref="SWISS-PROT:Q9BTC0"
    /translation="MDKQDPSENEAPPAIKPSEKPRKTRGRTTIARBEAGDAEADLPPEPQQQLGSLRSRQPKKTERVEQPLTIARRGRKMPVSLDSCEPTSCPATDERTASGVSASSTRSGPOSASTVVKRPPASBRVKGDDHDSDSDIGITLKELOHRLARKKEQSEPTERPLKGIQRLRKRRBEGPAPVGSASDITVEGVLPK"
    39..561
    /note="Cpg island"
    /evidence=not experimental
    complement(1..426..650,3246..3442,15475..15507))
    /gene="DATP1"
    /product="d5563B14.1 (Death associated transcription factor 1 (contains KIAA0333) isoform 1)"
    /note="Death associated transcription factor 1 (translation of cDNA PLACE1009138 (Em:AK002127)) isoform 2 alternatively spliced in 5' UTR"
    match: ESTs: Em:AW410673"
    /evidence=not experimental
    complement(1177..1649)
    /gene="DATP1"
    /note="match: GSS: Em:AQ127293"
    1327..1417
    /note="MULTI repeat: matches 439..535 of consensus"
    1462..1514
    /note="AluSc repeat: matches 246..298 of consensus"
    1749..2041
    /note="match: GSS: Em:AQ253731"

```

<http://www.sanger.ac.uk/BGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RP4-563B14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP4-305P22 is at 28420 in this sequence. The true right end of clone RP5-885L7 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP4-563B14 is from the library RP4-4 constructed by the group of Pliet de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

repeat_region	1786. .2093	/note="AluSq repeat: matches 1. .313 of consensus"
repeat_region	2898. .2953	/note="MER2 repeat: matches 97. .153 of consensus"
misc_feature		Complement(12912. .3305)
		/gene="DAP1"
repeat_region	3030. .3097	/note="match: STS: Em:N55493"
repeat_region	4231. .4366	/note="MER2 repeat: matches 269. .336 of consensus"
repeat_region	4520. .4828	/note="AluDb repeat: matches 3. .145 of consensus"
repeat_region	5459. .6077	/note="AluDb repeat: matches 1. .302 of consensus"
repeat_region	6078. .6366	/note="Char1a1 repeat: matches 2193. .2759 of consensus"
repeat_region	6367. .6994	/note="AluDb repeat: matches 1. .266 of consensus"
repeat_region	7348. .7470	/note="Char1a1 repeat: matches 4. .2193 of consensus"
repeat_region	8302. .8594	/note="AluDb repeat: matches 136. .259 of consensus"
repeat_region	9647. .9932	/note="AluSq repeat: matches 1. .301 of consensus"
repeat_region	9933. .10231	/note="Char1a1b repeat: matches 2. .239 of consensus"
repeat_region	10232. .10392	/note="AluSq repeat: matches 1. .300 of consensus"
repeat_region	10669. .10836	/note="Char1a1b repeat: matches 239. .523 of consensus"
repeat_region	10781. .10901	/note="AluDb repeat: matches 135. .301 of consensus"
misc_feature	10898. .11020	/note="match: STS: Em:G32962"
misc_feature		/gene="DAP1"
repeat_region	10956. .11135	/note="match: GSS: Em:AQ387275"
repeat_region	11138. .11305	/note="3 copies 60 mer 78% conserved"
repeat_region	11146. .11309	/note="14 copies 12 mer 67% conserved"
misc_feature		/note="41 copies 4 mer atac 64% conserved"
misc_feature		Complement(11548. .12516)
repeat_region	12058. .12208	/note="match: GSS: Em:AQ415373"
repeat_region	12213. .12509	/note="PRAM repeat: matches 1. .144 of consensus"
misc_feature	12555. .13045	/note="AluSq repeat: matches 1. .301 of consensus"
misc_feature	13880. .14299	/note="match: GSS: Em:AQ804060"
repeat_region	14246. .14317	/note="match: GSS: Em:AQ132474"
misc_feature	14712. .15898	/note="L2 repeat: matches 2581. .2658 of consensus"
misc_feature		/note="CpG island"
misc_feature		evidence=not experimental
repeat_region	15283. .15310	/gene="DAP1"
repeat_region	15515. .15636	/note="match: GSS: Em:AF190856"
misc_feature		/note="2 copies 61 mer 86% conserved"
misc_feature	18118. .18638	/note="CpG island"
misc_feature		evidence=not experimental
repeat_region	18989. .19268	Complement(18967. .19431)
repeat_region		/gene="DAP1"
repeat_region		/note="match: GSS: Em:AQ637346"
misc feature		/note="AluSq repeat: matches 2. .312 of consensus"
misc feature		Complement(18956. .19482)

	/gene="DATP1"	
	/note="match: GSS: Em:AQ764765"	
repeat_region	19976. .20270	
	/note="AluNb repeat: matches 1. .302 of consensus"	
misc_feature	complement(20348..20659)	
	/gene="DATP1"	
	/note="match: GSS: Em:AQ46936"	
repeat_region	21084. .21205	
	/note="FLAM_C repeat: matches 6. .129 of consensus"	
repeat_region	21248. .21364	
	/note="AluNb repeat: matches 2. .119 of consensus"	
unsure	21547.	
repeat_region	21967. .22272	
	/note="AluSp repeat: matches 1. .308 of consensus"	
repeat_region	22632. .22881	
	/note="AluNb repeat: matches 9. .263 of consensus"	
repeat_region	23208. .23574	
	/note="L2 repeat: matches 1520. .1917 of consensus"	
repeat_region	24500. .24660	
	/note="L2 repeat: matches 2372. .2542 of consensus"	
repeat_region	24758. .25032	
Query Match	25.1%;	Score 654.6; DB:9;
Best Local Similarity	98.7%;	Pred.No.2.7e-164;
Mismatches	0;	Mismatches 9; Indels 0; Gaps 0;

Query Match	25.1%	Score 654.6	DB 9	Length 28519
Best Local Similarity	98.7%	Pred. No. 2.7e-164		
Matches 660	Conservative 0	Mismatch 9	Indels 0	Gaps 0
QY 244	GCTTTCAGTGTGAGAGAGTATGAGAGACAAAGAGCGACCCGACCAATGAGAGGACACT			303
Db 669	GCTTACCTTACCGCTAGGTATGAGACAAAGGCGACCCGACCAATGAGAGGACACT			610
QY 304	AAGGCATCAAAACCCACACGACAAAGAGTTCCAGAAAAATGGGGTCTTTGAGAGACACT			363
Db 609	AAGGCATCAAAACCCACACGACAAAGAGTTCAAGAAAAATGGGGTCTTTGAGAGACACT			550
QY 364	ATCGCCAAAGCAGAGGGGCGCAGAGGGAGCGCGAGGCTGACCCACTGAGAGCGCCACCCCA			423
Db 549	ATCGCCAAAGCAGAGGGGCGCAGAGGGAGCGCGAGGCTGACCCACTGAGAGCGCCACCCCA			490
QY 424	CAGCAGCAGCTGGGGACTGTCCCTCGCGGCGCAGTGGAGAGCAGCCCAAGCGCACTGAGCGC			483
Db 489	CAGCAGCAGCTGGGGACTGTCCCTCGCGGCGCAGTGGAGAGCAGCCCAAGCGCACTGAGCGC			430
QY 484	GTGAGCAGTTCCTGACCATTTGCGCGCGCGCGCGCAGAGAGACATGCTGTCTCCCTG			543
Db 429	GTGAGCAGTTCCTGACCATTTGCGCGCGCGCGCGCAGAGAGACATGCTGTCTCCCTG			370
QY 544	GAGGATTTCTGTGAGCCCAAGTCTCTGCCCGCGCACAGACGCGGAGACAGCTCCGAGGGC			603
Db 369	GAGGATTTCTGTGAGCCCAAGTCTCTGCCCGCGCACAGACGCGGAGACAGCTCCGAGGGC			310
QY 604	AGCGTGAAGAAACGCTTCTGAGAGACAGAAAGCGAGCCCGCAGTCTGCTTCCACAGCTGTGAAG			663
Db 309	AGCGTGAAGAAACGCTTCTGAGAGACAGAAAGCGAGCCCGCAGTCTGCTTCCACAGCTGTGAAG			250
QY 664	GAACGACCAAGCTCTTCTGAAAAAGTGAAGAGAGGGAGTGAACAAGATGACACTCCGAT			723
Db 249	GAACGACCAAGCTCTTCTGAAAAAGTGAAGAGAGGGAGTGAACAAGATGACACTCCGAT			190
QY 724	AGTGAACAGCAATGCGCTGACCTTGAAGAAAGCTTCAAGAAATGCGCTTCCGACGAAACCGGAA			783
Db 189	AGTGAACAGCAATGCGCTGACCTTGAAGAAAGCTTCAAGAAATGCGCTTCCGACGAAACCGGAA			130
QY 784	CAGGAGCCCACTGAGAGGCGCCCTGAAAAAGGATTCAGAGTGCCTCTGCGGAAACCGCGG			843
Db 129	CAGGAGCCCACTGAGAGGCGCCCTGAAAAAGGATTCAGAGTGCCTCTGCGGAAACCGCGG			70
QY 844	GAGAGAGGATCCCGCCAGACTGTGGGCTCCGAGGCCAGTGAACACTGTGAGAGGCGTCTTG			903
Db 69	GAGAGAGGATCCCGCCAGACTGTGGGCTCCGAGGCCAGTGAACACTGTGAGAGGCGTCTTG			10
QY 904	CCCACTGAAG 912			
Db 9	CCCACTGAAG 1			

. Tue Apr 27 10:09:37 2004

us-09-787-016a-1.rge

Page 25

Search completed: April 23, 2004, 14:16:14
Job time : 9971.23 secs

CC Inducer-oblierator 1 (DIO-1) polypeptides. The polypeptides can be
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC agonists and antagonists are used as a medicament for treating diseases
CC characterized by an alteration in cell death or by hyperproliferation,
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC tumours, malignant tumours or hyperproliferative skin disorders. They are
CC also useful in the treatment of metabolic, proliferative or inflammatory
CC conditions. The present sequence represents a cDNA encoding the murine
CC DIO-1 polypeptide
CC
XX

SQ Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;

Query Match 100.0%; Score 2867; DB 3; Length 2867;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2867; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TCCGTGTAGCTCTGGAATAGCTCCGGGATCCGGCCGCGGAGCTTTTAAAGG 60
DB 1 TCCGTGTAGCTCTGGAATAGCTCCGGGATCCGGCCGCGGAGCTTTTAAAGG 60
QY 61 CAGTCCCACTGTCTTGTGTGTCAGAGCTCCGAATCTTCTCATTTATGAACTGTG 120
DB 61 CAGTCCCACTGTCTTGTGTGTCAGAGCTCCGAATCTTCTCATTTATGAACTGTG 120
QY 121 ACTGACAAAGTGGGGTGAAGGCTTGGCGTCTGACCTGACCTGACCTATATATTT 180
DB 121 ACTGACAAAGTGGGGTGAAGGCTTGGCGTCTGACCTGACCTATATATTT 180
QY 181 TATGTAGATTTCCAGCCAAAGTTTCCAGCTTCAAGTTTGGGACAGATATGATAT 240
DB 181 TATGTAGATTTCCAGCCAAAGTTTCCAGCTTCAAGTTTGGGACAGATATGATAT 240
QY 241 AAAGGCACTGAGACATAGAGAGACACCAAGCTTATCAACCACTGATAGAGATT 300
DB 241 AAAGGCACTGAGACATAGAGAGACACCAAGCTTATCAACCACTGATAGAGATT 300
QY 301 AGGAAACCTGGGGTTTGAAGAACCAATGTCGAAAGTGAAGGTGACAGAGACAG 360
DB 301 AGGAAACCTGGGGTTTGAAGAACCAATGTCGAAAGTGAAGGTGACAGAGACAG 360
QY 361 GAGGCGGACCCGAGTGAAGACCAAGAGCATTAACCTTCCCTGCGCGAGTGA 420
DB 361 GAGGCGGACCCGAGTGAAGACCAAGAGCATTAACCTTCCCTGCGCGAGTGA 420
QY 421 CGGCAACCAAAAGCTGAGAGGTTAGAGATTCTTACACCGCTTCCGCGAGG 480
DB 421 CGGCAACCAAAAGCTGAGAGGTTAGAGATTCTTACACCGCTTCCGCGAGG 480
QY 481 AAAAAAGATGTGCGGTGCTGAGAGATTCAGTGAAGCCACATCTTCCACAGTCA 540
DB 481 AAAAAAGATGTGCGGTGCTGAGAGATTCAGTGAAGCCACATCTTCCACAGTCA 540
QY 541 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAAGTTCTGACATGAAAGTGC 600
DB 541 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAAGTTCTGACATGAAAGTGC 600
QY 601 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAAGTTCTGACATGAAAGTGC 660
DB 601 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAAGTTCTGACATGAAAGTGC 660
QY 661 GAGGAAAGAACACCTGACAGTGAAGTGAAGCTTGAAGGAACTTCAGAAC 720
DB 661 GAGGAAAGAACACCTGACAGTGAAGTGAAGCTTGAAGGAACTTCAGAAC 720
QY 721 CGCCTTCCGAGAAACGAGAGCAAACTGTGAGAGGTCCTGAGAGGAGTCAAA 780
DB 721 CGCCTTCCGAGAAACGAGAGCAAACTGTGAGAGGTCCTGAGAGGAGTCAAA 780
QY 781 CGCCTGAGAGAAACGAGAGCAAACTGTGAGAGGTCCTGAGAGGAGTCAAA 840
DB 781 CGCCTGAGAGAAACGAGAGCAAACTGTGAGAGGTCCTGAGAGGAGTCAAA 840
QY 841 AGTCCGAGAGCAAGACCTCTGTGAAGCAGAGCTGAGGCTTGAAGGGAACA 900
```

```
DB 841 AGTCCGAGAGCAAGACCTCTGTGAAGCAGAGGCTGAGGCTTGAAGGGAACA 900
QY 901 TCCGAGTGAAGACAGATGACATGAAATCATGTTGAAGGAGGCACTCAAGGAA 960
DB 901 TCCGAGTGAAGACAGATGACATGAAATCATGTTGAAGGAGGCACTCAAGGAA 960
QY 961 ACAGAGAAACCCGAGGAAAGCGGGCAACCAAGCTGAGTGAAGGTTTACAGACC 1020
DB 961 ACAGAGAAACCCGAGGAAAGCGGGCAACCAAGCTGAGTGAAGGTTTACAGACC 1020
QY 1021 AATGCCCTGTACTGATCTGACCTGCGGCAAGCTTCAACCAAGGTTATGATCTG 1080
DB 1021 AATGCCCTGTACTGATCTGACCTGCGGCAAGCTTCAACCAAGGTTATGATCTG 1080
QY 1081 CGGTGTGAGAGAGTGTTCATGAGTGAAGTGTGAGTATTTCTGAGGCGCGGCT 1140
DB 1081 CGGTGTGAGAGAGTGTTCATGAGTGAAGTGTGAGTATTTCTGAGGCGCGGCT 1140
QY 1141 CTGGAAGAAACGAGGAAAGCACTGACCTGACCAATTCACCAATTTGCAAGTGA 1200
DB 1141 CTGGAAGAAACGAGGAAAGCACTGACCTGACCAATTCACCAATTTGCAAGTGA 1200
QY 1201 GAGCAAAACGAGTGAAGCCCAATGAGAGAGCTCTGAGTGCAGATCTGTGCTGAT 1260
DB 1201 GAGCAAAACGAGTGAAGCCCAATGAGAGAGCTCTGAGTGCAGATCTGTGCTGAT 1260
QY 1261 GGCACAGACTGCAAGATGAGGACATGAGAGCAAGTCCGAGAAACCAAGGCA 1320
DB 1261 GGCACAGACTGCAAGATGAGGACATGAGAGCAAGTCCGAGAAACCAAGGCA 1320
QY 1321 AAGGTAAGATTTGAAGAGGAGCAAAACCCAGCGGCAAGAAATCAAGATATTCAG 1380
DB 1321 AAGGTAAGATTTGAAGAGGAGCAAAACCCAGCGGCAAGAAATCAAGATATTCAG 1380
QY 1381 CCTGTGTAGAGGCTCTGTGCTCTTAATGCAATGACCTGAGGTTCAGTGA 1440
DB 1381 CCTGTGTAGAGGCTCTGTGCTCTTAATGCAATGACCTGAGGTTCAGTGA 1440
QY 1441 CAGCTGACTGTGTATTTGAGATGATGATGATCTTCAACAGGAGAGCTTACAT 1500
DB 1441 CAGCTGACTGTGTATTTGAGATGATGATGATCTTCAACAGGAGAGCTTACAT 1500
QY 1501 AGATTTCTAAGTTCAGGTAAGAACCAAAACCAAGGAAAGGTCAAGAGAG 1560
DB 1501 AGATTTCTAAGTTCAGGTAAGAACCAAAACCAAGGAAAGGTCAAGAGAG 1560
QY 1561 CCAAGAAAGTTCAAGTCTTCAAAATGAGTTCAGTGAAGTTAAATCTTCTGTG 1620
DB 1561 CCAAGAAAGTTCAAGTCTTCAAAATGAGTTCAGTGAAGTTAAATCTTCTGTG 1620
QY 1621 CACAGAGACATGAGCTTCAAGAGAAAGGAAACCAAGTGAAGTGTGCTTCC 1680
DB 1621 CACAGAGACATGAGCTTCAAGAGAAAGGAAACCAAGTGAAGTGTGCTTCC 1680
QY 1681 AGGATGAGACCTTCTGAGAGAGGAGCAGCTGTGAGAGCAAGACCAATCTGAG 1740
DB 1681 AGGATGAGACCTTCTGAGAGAGGAGCAGCTGTGAGAGCAAGACCAATCTGAG 1740
QY 1741 GACCAAACTACATGCTGTGAGAGCAGAGAGCAAGAGCCCACTGACCTTCC 1800
DB 1741 GACCAAACTACATGCTGTGAGAGCAGAGAGCAAGAGCCCACTGACCTTCC 1800
QY 1801 ACCCTATTGAGTAAATGTAAGTAAACCAAAAGGCTTCCAGGCGGCTCCAT 1860
DB 1801 ACCCTATTGAGTAAATGTAAGTAAACCAAAAGGCTTCCAGGCGGCTCCAT 1860
QY 1861 CTGGGTGCTGCTGAGGCTGTCTAGAGCAGAGTCTGAGGTTCTGAGTGTGAT 1920
DB 1861 CTGGGTGCTGCTGAGGCTGTCTAGAGCAGAGTCTGAGGTTCTGAGTGTGAT 1920
QY 1921 GCGAGAGCTCACTGACGAGAGCAAGATCAAGATGCTTGAAGCCAGGTTTC 1980
```

```

Db      1921 GCCAGCAGCTGACGCGCAGCCAGAGATACCAAGATGCTCTGAGCCCCAGGTGTC 1980
Qy      1981 CTGCTAGCCTGTGGAGGCTCTCTGGGTGTTCTTAAAGAGCTGTAGGCTCATGTG 2040
Db      1981 CTGCTAGCCTGTGGAGGCTCTCTGGGTGTTCTTAAAGAGCTGTAGGCTCATGTG 2040
Qy      2041 GAGGCAATTTCTTATTTAGGCTTGTGATTTGAACTGTCTCAACAGCA 2100
Db      2041 GAGGCAATTTCTTATTTAGGCTTGTGATTTGAACTGTCTCAACAGCA 2100
Qy      2101 TGGGCAACATCAGTGCCTGTTGAGCTCAGCTTGAAGTTCCTCAGAGCAGC 2160
Db      2101 TGGGCAACATCAGTGCCTGTTGAGCTCAGCTTGAAGTTCCTCAGAGCAGC 2160
Qy      2161 CCAGCACAACCATGAGATTTTCCACGTTACCTATTTTCTTCAACATGTCACCACTG 2220
Db      2161 CCAGCACAACCATGAGATTTTCCACGTTACCTATTTTCTTCAACATGTCACCACTG 2220
Qy      2221 CTCTTCTATAGAAAAGTATTTTTCATGAGTTAGGCAATTTTGAAGCAGAGTAGA 2280
Db      2221 CTCTTCTATAGAAAAGTATTTTTCATGAGTTAGGCAATTTTGAAGCAGAGTAGA 2280
Qy      2281 CAATTTGTCTATTTTCACTACTACACCTTACATTAACAGGTTGATTTTGAAGTTC 2340
Db      2281 CAATTTGTCTATTTTCACTACTACACCTTACATTAACAGGTTGATTTTGAAGTTC 2340
Qy      2341 ATCTGCTGCTAGATCCATGCTGATTAATGCTGACATACACCTTACAGCAGATTC 2400
Db      2341 ATCTGCTGCTAGATCCATGCTGATTAATGCTGACATACACCTTACAGCAGATTC 2400
Qy      2401 TCATCAAAAGTGAATGTGTGAGTGTGCTTCCAGCTCTGATAGTCCCAAGATTTC 2460
Db      2401 TCATCAAAAGTGAATGTGTGAGTGTGCTTCCAGCTCTGATAGTCCCAAGATTTC 2460
Qy      2461 TCAGTGTATTAGTTGAGTAACTTTTGAACCCAGATTGATCACTGAGTTGGAGC 2520
Db      2461 TCAGTGTATTAGTTGAGTAACTTTTGAACCCAGATTGATCACTGAGTTGGAGC 2520
Qy      2521 CATCTGTTTCACTGCTGTACAGGATCACTTTCTTACTTTTGAAGTGTTC 2580
Db      2521 CATCTGTTTCACTGCTGTACAGGATCACTTTCTTACTTTTGAAGTGTTC 2580
Qy      2581 GTGAATTTTCACTGCTGTACAGGATTTTGTTCATCTCTGTCACCTCAGCC 2640
Db      2581 GTGAATTTTCACTGCTGTACAGGATTTTGTTCATCTCTGTCACCTCAGCC 2640
Qy      2641 CCAGATTAAGGGGCTCATGCTTAAAGCTGCTTAAATTTTCTGTAAGTTGGGGTCC 2700
Db      2641 CCAGATTAAGGGGCTCATGCTTAAAGCTGCTTAAATTTTCTGTAAGTTGGGGTCC 2700
Qy      2701 TCCCTGCCCCCATCTGTAAGAGTGTGTAACCTTACAGGACACACTGATGATGCTGTG 2760
Db      2701 TCCCTGCCCCCATCTGTAAGAGTGTGTAACCTTACAGGACACACTGATGATGCTGTG 2760
Qy      2761 GGGTGTGAATCAGAGTGTGACAGCAGCTTACAGCTGAGCTGATTAAGCCATG 2820
Db      2761 GGGTGTGAATCAGAGTGTGACAGCAGCTTACAGCTGAGCTGATTAAGCCATG 2820
Qy      2821 GTTACTGAATCTGATCTTCTTGTCTTTGTTTCTTTGTGCTGGG 2867
Db      2821 GTTACTGAATCTGATCTTCTTGTCTTTGTTTCTTTGTGCTGGG 2867

```

```

RESULT 2
AA290578 standard; cDNA; 2610 BP.
AA290578;
AA290578;
19-JUN-2000 (first entry)
Human death inducer-obliterato 1 (DIO-1) polypeptide encoding cDNA.

```

```

KW      Death inducer-obliterato 1; DIO-1; cell death; cancer; tumour; human;
KW      autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW      antirheumatic; antiinflammatory; antiproliferative; ss.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      CDS      265..1953
FT      /tag=a
FT      /product="DIO-1"
PN      MO200015787-A1.
PD      23-MAR-2000.
PF      10-SEP-1999; 99WO-GB003019.
PR      10-SEP-1998; 98SR-00003069.
PR      17-SEP-1998; 98US-0100873P.
PA      (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA      (BANU/) BANNERMAN D G.
PI      Alonso CM, Domingo DG, Grandien A, Leonardo B, Martinez P;
DR      WPI; 2000-271426/23.
DR      P-PSDB; AAY67579.
PT      New DNA encoding human and murine death inducer-obliterato 1
PT      polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT      diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT      hyperproliferative skin disorders.
PS      Claim 1; Fig 1A; 27pp; English.
CC      The invention provides nucleic acids encoding the human and murine death
CC      inducer-obliterato 1 (DIO-1) polypeptides. The polypeptides can be
CC      expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC      agonists and antagonists are used as a medicament for treating diseases
CC      characterized by an alteration in cell death or by hyperproliferation,
CC      e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC      tumors, malignant tumors or hyperproliferative skin disorders. They are
CC      also useful in the treatment of metabolic, proliferative or inflammatory
CC      conditions. The present sequence represents a cDNA encoding the human DIO
CC      -1 polypeptide
SQ      Sequence 2610 BP; 644 A; 667 C; 734 G; 565 T; 0 U; 0 Other:
Query Match      35.7%; Score 1023.6; DB 3; Length 2610;
Best Local Similarity 71.9%; Pred. No. 3.3e-309;
Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;
Qy      1 TCCGTGTAGCTCTGGAATGCTGTGGGATCCCGCGCCGGGAGCTGTGTTAAGG 60
Db      21 TCCGTGTAGCTCTGGAATGCTGTGGGATCCCGCGCCGGGAGCTGTGTTAAGG 80
Qy      61 CAGTCCCACTGCTCTGTTGTGTTGCAAGCTCCGGAATCTTCTATGATGAATG 120
Db      61 CAGTCCCACTGCTCTGTTGTGTTGCAAGCTCCGGAATCTTCTATGATGAATG 140
Qy      81 GAAACGCTCTGATTAATCTGATGTTGTAATAATACAAAGCTGTTACTGTGAACAG 160
Db      81 GAAACGCTCTGATTAATCTGATGTTGTAATAATACAAAGCTGTTACTGTGAACAG 180
Qy      121 ACTGACAAAGTGGGGTGA--GCTTGGCCGTCTGCTGAC-----TGGCC 166
Db      141 GGTGACAAAGTGTGTTGTGAGCTGGGTGCTGTGCTGGAACCAAGGTTGTGTCG 200
Qy      167 CAGGTCTAATTTTATGAGATTTCCAGCCAAAGTTCACAGTTTCAATTTTGGGA 226
Db      201 AGGTTTTTGTGTATTTAGATTTT-CAGGGAAGTGTCCAAAGTTCAGTGTGGAG 259
Qy      227 CAGGTATGATGATTAAGGCACTGAGCAATGAGGACCAACCAAGGCTATCAAAACCA 286
Db      260 CAGGTATGATGATTAAGGCACTGAGCAATGAGGACCAACCAAGGCTATCAAAACCA 319
Qy      287 CCAGTAAAGATTCAGGAAACCTGGGGTTTTTGAAGAAACCAAGATTGCCAAAGCTGAGG 346

```

```
Db 320 CCAGCAAGAGTTCAGAAAACATGGGGTTTTCAGAGACCATATCCCAAGCAAGAG 379
Qy 347 GTGACAGAGACACGAGGCGGACCCCACTGAGCAGCAAC-----ACAGCAGCATTAAC 400
Db 380 GCGCAGGGGAGAGCGGAGAGCTGACCCCACTGAGGCGGCAACCCCAAGCAGCAGCTGGGCC 439
Qy 401 TCTCCCTGCGCGCAGTGAAGGCAACCAAAAGTATGAGAGGCTAGAGAGTTTCTTA 460
Db 440 TGTCCCTGCGCGCAGTGAAGGCAAGCCCAAGCAGCATGAGCGCGCTGAGGAGTTCTGA 499
Qy 461 CCAAGCTTGGCGCGCAGGAGAAAAGATGCGCGGCTGCTCCCTGAGAGATTCAGTAGGC 520
Db 500 CCAATTCGCGCGCGCGCGCAGAGAGACATGCTGTCTCCCTGAGAGATTCGTGTAGGC 559
Qy 521 CCAATCTTCAACAGTCACTGATGTGAGACAGCTTCCGAGGGGAGCGTTGAAAGCACTT 580
Db 560 CCAACTCTGCGCGCGCGCAGACCGCGAGACAGCTTCCGAGGGGAGCGTTGAAAGCGCTT 619
Qy 581 CTGAGATCAGAAAGTGGCCCTGTATCTGACTCTTA--GGGAAAGACATCTGCTCTT 637
Db 620 CTGAGACAGAAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAAGAGACAGCAGCTCTT 679
Qy 638 CTGAAAAGGCAAAAGAGGTGAAGAGAAAGACACCTGACAGTGAAGTGAAGTGGCC 697
Db 680 CTGAAAAGGTGAAGAGAGAGAGTACACAGATGACCTCCGATAGTGAAGAGATGGCC 739
Qy 698 TTACGTTGAAGAACTTCAGAACCCCTTCGAGAAAGGAGAGAGCAAGCACTGTGAGA 757
Db 740 TGACCTTGAAGAAAGCTTCAGAAATGCCCTTCGAGAAAGGAGAGAGAGCCCACTGAGA 799
Qy 758 GGTCCCTGAGAGGCAAGTGAATGCCCTGAGAGAAAGTGCAGAGAGAGAAATTCGCCG 817
Db 800 GGGCCCTGAAAGGGATCCAGAGTCCCTGCGGAGAAAGCGCGAGAGGGGTCCCGCG 859
Qy 818 AAACGTGGAGAGTCCAAATAGGCAAGTCCGAGCAGACAGACCTCTGTATAGCAGAGC 877
Db 860 AGACTGTGGGCTCCAGGCGCAGTGAACACTGTGAGAGGCGCTCCGCGCAGTATAGCAGAGC 919
Qy 878 CTGAGGCTAGTCAAGGAGCAGATGTCCAGTCAAGACAGATGACATAGAAAATCAATGG 937
Db 920 CCGAAGACGATCAGAGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGATTAAGTGG 979
Qy 938 AAGGAGAGGCGCATGAGGGAATTAAGAGAAAACCCAGGAGAGCGGCAAAACCAAGC 997
Db 980 AGGAGAAAGCGGCTCAGAGACATCAAAAGATGAGAGACCTTGAGACTTGGCCGACGAAAGC 1039
Qy 998 CTGAGTGTAGAGTTTACGACCCCAATGCGCTGTACTGCACTGCGGCGCAGCTCAACA 1057
Db 1040 CTGAATGTAGAGGTTTACGACCCCAAGCGCTGTATTTGCGCGCAGCTCAACMACA 1099
Qy 1058 ACAGGTTATGATCTGCTGTGATCCGATGAGAGTGGTTCATGAGTGTGGTGA 1117
Db 1100 ACAGGTTATGATTTGCTGTGACCGCTGTGAAGAAATGGTTTCAATGGCGATGTGTGGGA 1159
Qy 1118 TTTCTGAGGCGCGGAGCGGCTCTGTGAAGGAAGGGGAGAGATTAATCTGCGCAATT 1177
Db 1160 TTTCTGAGGCTCGAGGAGGCTTTTGTGAAGAAAGGGGAGAACTATATCTGCGCAACT 1219
Qy 1178 GCACCATTTTGAAGTGAAGATGAGACAAACGATAGCGCCCAATGAGCAGAGACTGTG 1237
Db 1220 GCACCATTTCTGCAATGTGAGATGAGATCACTTTCAGAAAAGCAGATCAGCAGAGAGCTA 1279
Qy 1238 GGTGCAATCTGTGGGTCTGTATGACACAGACTGACAGATGAGGAGCAGTATGAGCAGA 1297
Db 1280 AATGAGACCTGAGAGATGTGTATGACACGATTTGTACAAAGTATAGAAACAAATAGAGCAGA 1339
Qy 1298 AGTCCGAGAAAGACCAAGGCGCATTAAGGGTATGAGATTTGAAGAGCAGCAAAACCCACGCGA 1357
Db 1340 AGTCTAGGAAAGACCAAGGAGTAAAGGGTATGAGATTTGAAGAACTCAATTCAGATGGCA 1399
Qy 1358 AGAAAACTCAAGATATTTCCAGCTGTGTGAGAGGCTCTGTAGATCAATTTG 1417
```

```
Db 1400 AGAAGAACTCAAGATCTTCCAGCCTGTGTATGATAGAGCGCGCTGTGCTCAAAATGTATTG 1459
Qy 1418 GCCCTGGGTGTTCCAGTGTAGACAGCCTGACTCTGTGTATTTGCAGTATGACTGTCAATTC 1477
Db 1460 GCCCGGGTGTCTGTACCTGTGGCGGAGCCGACTCGGTGTATCTGCAGTATGACTGTATTC 1519
Qy 1478 TCAAACAGGCGAGCTTCAATGAGATTTCTTAAGTTCAAGTTAAAGAACAAAACAAAC 1537
Db 1520 TCAAACAGCGCGCAGCGCAATTAAGTTTCTTAAGCTCAAGTTAAAGAACAAAGCCAAAGC 1579
Qy 1538 CCAAGAAAAGGTCAAGACGAAAGCCAGAAAAGTTCACTTTCAAATATGACAGTGTTCAG 1597
Db 1580 CTAAAGAAAAGATGAAGATTAAGACCAAGAGAGCCAGTCTTCGAAAATGGCGGTCTCAG 1639
Qy 1598 TGGGATTAATAATCTTTCTGTGACAAAGAGCTAGCGTCAAGAAAGAGGAAACCCAG 1657
Db 1640 CAGGTATTAATAATCTTTCTGTGACAAAGAGCAGCTTCAGAAAAGAGAGACCAAG 1699
Qy 1658 TGAAGAA---AGTATGTGGCTTCCAGAGTGAAGACTTCTGGAAAGGAGGAGCGCTGTG 1714
Db 1700 TGAAGAGGCAAGTGTGTCTTCCGCGAGATGAAGACTCGGAGAGAGAGAGAGCTGTG 1759
Qy 1715 AGAGCAGCACACATCTGTGGCAAGTGAACAACATAATGCTGTGAAGCCAGAGAGC 1774
Db 1760 AGAGCAGCACCGCGTGTGGGAGAGCATCAATTAACATGCAAGTAAAGCCAGAAAAGA 1819
Qy 1775 CAGAGAGCCCACTGCACTGTGCCACCCCTATTGATTAATATATATATATATATATATAT 1834
Db 1820 CTG-----CTGCTCCCTGCGCGTCACTGTGTATTAATGTATATATATATATATATAT 1870
Qy 1835 CTGCTTCCAGGCGCGCTCCCATCATCT-----GGGTGGCTGCTGTGGGCTGTCTTAGAGC 1889
Db 1871 TTGGCTCTTGACCCCTCCGTTCTTTCTGATATAGCAATCCCTGTGGCTGTCTTAGAGC 1930
Qy 1890 CAG-----AGTCTGGGTGTCTGTGTGCTGATAGTAGCCAGAGC---TCACTGCCAGC 1940
Db 1931 TGGAGTTGCAAGCTTGTGTATTAAGCTGATCAAGACACCGGCTGACCATCAGCGGAG 1990
Qy 1941 CAGAGCAGATTAACAAATGTCTCTGAGACCCCAAGTGTTCCTAGCTGTGAGAGCT 2000
Db 1991 CAGAGCCCATGTCCAGATATGCTCTGTGCTCTGTGTGCATCCCTAGTGTGTCAAGACT 2050
Qy 2001 CTCTGGGTGTCTCTTAAGAGCTGTGAGGCTCATGT 2038
Db 2051 TCTGTCACTGTTTCCAAAGCTGAACCTCATGCT 2088

RESULT 3
AAH15102
ID AAH15102 standard; cDNA; 2772 BP.
XX
AC AAH15102;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human CDNA sequence SEQ ID NO:13129.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-0018776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
```


QY 1689 GACTTCTGGAGAGGAGGAGCCTGTGAGAGCAGCAACCATTCTGGGAGAGTACCA 1748
 DB 1770 AGCACTGGGAGAGGAGGAGGCTTGTGAGAGCAGCAGCCCTGTGGGAGAGGATCA 1829
 QY 1749 CTACATCTGTGAAGCCAGAGAGCCAGAGAAAGCCACTGACTCTGGCCACTTAT 1808
 DB 1830 TTACATCTGTGAAGCCAGAGAGCCAGAGAAAGCCACTGACTCTGGCCACTTAT 1880
 QY 1809 GAGTAAATGTAGTATCAACCCAGAGGCTTCCAGAGCCCTCCCATCATCT-----G 1863
 DB 1881 GTATTAATGTATGTATCACTTATAGGAGTGGCTCTCTGAGAGCCCTCCCTCTTCTGAT 1940
 QY 1864 GGTGCTCTGTGGGCTGTCTAGAGCAG-----AGTCTGGGTGTCTGTGTGATAG 1918
 DB 1941 AGCCATCCCTGTGGGCTGTCTAGAGCAGCTGGAGTTCAGCTTGTGTAGCTGATCAG 2000
 QY 1919 TAGCAGAGAGC-----TCACTGCAAGCCAGAGAGCATATCAAGATACCTTGGAGCC 1974
 DB 2001 ACACCGGCTGACCATCAGCCGAGAGAGAGCCCATGTCCAGATGCTCTGTGCTGCT 2060
 QY 1975 GGTGCTCTGTGGGCTGTCTAGAGCAGCTTCTGTGGGTCTCTTAAGAGCTGTAGGCTC 2034
 DB 2061 GTGTCAATCCCTGTAGTGTCTAGAGACTTCTGTCTCTTCTCAAGCTGTAAACCTAC 2120
 QY 2035 ATGT 2038
 DB 2121 TGGT 2124
 RESULT 4
 ABV28029
 ID ABV28029 standard; cDNA; 2085 BP.
 AC ABV28029;
 XX 16-SEP-2002 (first entry)
 DT Human prostate expression marker cDNA 28020.
 DE Human prostate expression marker cDNA 28020.
 XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 OS Homo sapiens.
 PN WO200160860-A2.
 XX 23-AUG-2001.
 PF 20-FEB-2001; 2001MO-US005171.
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Schlegel R, Endege WO, Monahan JE;
 DR WPI, 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX Claim 1; Page 5788; 11750bp; English.
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 CC
 SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;
 Query Match 33.1%; Score 949.8; DB 5; Length 2085;
 Best Local Similarity 73.5%; Pred. No. 46-286;
 Matches 1884; Conservative 0; Mismatches 462; Indels 37; Gaps 12;
 1 TCCGTGTAGTCTGTGAATATGCTGCGGATCCCGGCGCGGAGCTTGTTAAGAG 60
 DB 35 TCCGCGGCGTTCGGGAAATGCTGAGAGCCCTTGAAGGCTGCGGAGCTTCA 94
 QY 61 CAGTCCCACTGTCCCTTGTGTTCAGAGCTCCGAAATCTTCTCATTTAGACTGTG 120
 DB 95 GAACAGCCTCTATATATCTGAGTGTGAATAATAC-GAAGCTGTATCTGTGAACAGTG 153
 QY 121 ACTGACACAGTGGGT--GAGGCTTGGCCCTGTGCTT-----ACCTGGCCCC 167
 DB 154 GCTGACAAAGTGTGTGTGTGAGCTGTGCTGTCTGTGAACCCAGAGTTTGTCTGCC 213
 QY 168 AGGTCTATATTTTATATATATATTCAGCCCAAAGTTTCCAAAGTTTCAATTTTGGAG 227
 DB 214 AGGTTTGTGTGTATTTATATATATTTT-CAGGAAAGTGTCCAAAGCTTCAAG-1GTGAGC 271
 QY 228 AGGTATGATGATTAAGAGGACCTTGAAGATGAGAGAGCCCAAGCTTATCAAACCCAC 287
 DB 272 AGGTATGAGCAGCAAAAGGCGAGCCGAGCATGAGAGGAGCATTAAGGCCATCAAAACCCAC 331
 QY 288 CAGTAAGAGATTCAGAGAAACCTGAGGTTTTCAGAGAAACAGATGTCGCAAGAGTGAAG 347
 DB 332 CAGCAAAAGTTCAGAGAAACCTGAGGTTTTCAGAGAAACAGATGTCGCAAGAGTGAAG 391
 QY 348 TGCAGAGACACGAGAGGCGAGCCCAAGTGTAGAGAGCA-----ACCAAGAGAGATTAACCTC 402
 DB 392 CGCAGGAGAGCGAGAGGCTGACCTCACTGAGCGGAGCCCAAGAGAGAGCTGAGCCCTG 451
 QY 403 TCCCTGCGCCGAGTGTGACGCGCAACCAACCTACTGAGAGGTTGAAGATTTCTTACC 462
 DB 452 TCCCTGCGCGGAGTGTGAGGAGGAGCCCAAGCGCACCTGAGCGGAGTGTGAGAGCTTCTGACC 511
 QY 463 AGGTTGCGCGCGGAGGAGAAAGATGTCGCGGTGCTCTGAGAGATTCAGTGAAGCC 522
 DB 512 ATTGCGCGCGCGCGGAGGAGAGCATGCTGTCTCTGAGAGATTCAGTGAAGCC 571
 QY 523 ACATCTTCCACAGTCACTGATGTGAGACAGCTTCCGAGGAGAGCTTGAAGACATTTCT 582
 DB 572 AGCTCTGCGCGG-CACAGAGCGGAGAGCAGC-TCCGAGGCGAGCGTGAAGAGCGCTTCT 629
 QY 583 GAGATCAGAGTGGCGCTGTATCTGATCTCTTGAAGAAAGACATCTGCTCTTCTTGA 642
 DB 630 GAGACAGAGAGGCGCGCCAGTCTGTCCAGTGTGAAGAGAGCAAGCTTCTTGA 689
 QY 643 AAGGCAAAAG 702
 DB 690 AAGGTGAAG 748
 QY 703 TTGAAG 762
 DB 749 TTGAAG 807
 QY 763 CTGAAG 822
 DB 808 CTGAAG 867
 QY 823 GGAAGTGTCAATATAGGAG 882

Dp	868	GTGGGCTCCGAGGCCAGTGCACCTGTGGAGGGCTCTCCGCCAGTAAAGCAGAGCCCGAG	927
Qy	883	GCTAGTCAGGGACCAAGTGTCCCAAGTCAGAGCAGATGACATTAAGAAAATCACTTTGGAAAGG	942
Dp	928	AACGATCAGGGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGATTAAGTTTGGACGGGA	987
Qy	943	AAAGCGACTCAGGGAATATACAAAGGAAAAACCCAGAGGAAGCGGGCAAAACCAAGCTGAG	1000
Dp	988	AAAGCGGCTCAGGACATCAAAATATAGAGGCTTGAGAACTTATGGCCGACCGAAGCTTGAA	1047
Qy	1003	TGTGAGTTTACGACCCCAATGCCCTGTATCTCATCTGCGCCAGGCTCAACAACAAG	1067
Dp	1048	TGTGAGGTTACGACCCCAAGCCCTGTATTTGCAATTTGCCCGCAAGCTTCAACAACAAG	1100
Qy	1063	TTTATGATCTGCTGTGATCCGATGTGAGAGTGGTTCCATGTGACCTGTGTGGTATTTCT	1122
Dp	1108	TTTATGATTTGTGCTGTGACCGCTGTGAAGATGGTTTCAATGCGATGTGTGGGCACTTCTC	1167
Qy	1123	GAGGCCCGAGGGCGGCTCTGTGAAGGAAACGGGGAAAGATACATCTGCCAAATTGCAAC	1188
Dp	1168	GAGGCTCAGGGAGGCTTTTGGAAAGGAATGGGAAAGACTATATCTGCCCAATGCAAC	1222
Qy	1183	ATTTTGCAGGTGCAGAGTAGAGCAACAAGGTAGCGCCCAATAGAGCAGACTCTGGGTGC	1244
Dp	1228	ATTCTGCAGGTGCAGAGTAGAGCTCATTTCAAGAAACGGCAGATCAGCGAGAACTTAATGG	1288
Qy	1243	AGATCTGTGGGTCTGATGTGCACAGACTGCAACAACATTAAGGACAGTAGAGCAAGATCC	1300
Dp	1288	AGACTGTGAGATGTCTGATGTGCACCGCATTTGTATCAAGTAATAGAAACAATAGAGCAAGATCT	1344
Qy	1303	GGAAAGATCAGGGCATTAAGGTTAGATTAAGAGGAGCGCAAAACCCAGCGGCAAGAA	1366
Dp	1348	AGCGAAGATCAAGGGATTAAGGTTAGATTAAGAGGAGCTGCAAAATTCAGATGGCAAGAG	1400
Qy	1363	AAACTCAGATATTTCAAGCTGTCTGTAGAGGCTCTGTGGCTCTTAATATGACTTGGCCCT	1422
Dp	1408	AAACTCAGATATTTCAAGCTGTGTATAGAGGGGCTGTGGCTTCAAAATGTATTTGGCCCC	1466
Qy	1423	GGGTGTTCAGGTAGCACAAGCTGACTGTGTATTTGCAATTAATGACTCATTTCTCAAA	1488
Dp	1468	GGGTGTCTGTCAAGTGGCGCAGCCCGACTCGATGTACTGTGACATTAATGACTGTATCTCAAA	1522
Qy	1483	CACGACAGACTACCATAGATTTTCTTAAGTTCAAGTTAAAGAACAAAAACAAATCCCAAG	1544
Dp	1528	CACGCGCAGGACCAATGAAGTTTCTTAAGCTCAGGTAAAGAAACAGAAAGCAAAAGCCTTAA	1588
Qy	1543	GAAAAGGTCAAGACGAGCAGAAAAAGTTCACTTCTCCAAATATGCAAGTTTCAAGTGGGG	1600
Dp	1588	GAAAAGATGAAGATTAAGCCAGAGAAAGCCCACTTTCGAAATGGGGTGTCTAGCGAGGT	1644
Qy	1603	ATTAAATCTCTTCTGTGCAACAAGACTAGCGGTCAAGAAAAAGGAAAAACCAAGTAGAG	1666
Dp	1648	ATTAAATCTCTTCTGTGCAACAAGACCAAGCTCCAGAAAAAAAAGAACCAAGTAGAG	1700
Qy	1663	AA---AGTGAATGCTGGCTTCCAGAGAGTGAAGCTTCTGGGAAAGAGCAGGCTGTGAGAGC	1711
Dp	1708	AAAGCAGTGGTGGTCTCTGCGGGAGTGAAGCACTCGGAAAGAAAGCAGCTTGTGAGAGC	1766
Qy	1720	AGCAGCAGCATCTGGGGCAAGTAGACCAACAATCAATGTGTGTAAGAGCAGAGAGGACGAG	1777
Dp	1768	AGCAGCGGCTGTGGGCGAGGACATCAATTAATGCAAGTAAACCAAGAAAGACTG--	1822
Qy	1780	AAAGCCACTGCACTGTGGCCCAACCTTATTTAGTAAATGTAAGTATCAATCCAAAGGCTGGC	1833
Dp	1826	-----CTGCTCCCTCGCGCTGCACTGTGTGTAAATGTAATGATCACTAGGGTGGGC	1876
Qy	1840	TTCCAGGCGCCCTCCCAATCATCT	1862
Dp	1879	CTCCAGACCCCTCCGTTCTTT	1901

ID	ABV27935	ABV27935 standard; cDNA; 2085 BP.
XX	AC	ABV27935;
XX	DT	16-SBP-2002 (first entry)
XX	DE	Human prostate expression marker cDNA 27926.
XX	KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; pharmacogenomic marker; gene; ss.
XX	OS	Homo sapiens.
XX	FN	WO200160860-A2.
XX	PD	23-AUG-2001.
XX	PF	20-FEB-2001; 2001WO-US005171.
XX	PR	17-FEB-2000; 2000US-0183319P.
XX	PR	16-MAR-2000; 2000US-0189862P.
XX	PR	25-MAY-2000; 2000US-0207454P.
XX	PR	09-JUN-2000; 2000US-0211314P.
XX	PR	18-JUL-2000; 2000US-0219007P.
XX	PA	13-DEC-2000; 2000US-0235281P.
XX	PI	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	PI	Schlegel R, Endege WO, Monahan JB;
XX	DR	WPI; 2001-662795/76.
XX	PT	Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.
XX	PS	Claim 1; Page 5757-5758; 11750pp; English.
XX	CC	The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the efficacy of progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cancer cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX	XX	Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;
QY	Query Match	33.1%; Score 949.8; DB 5; Length 2085;
Db	Best Local Similarity	73.5%; Pred. No. 4e-286;
Db	Matches 1384; Conservative	0; Mismatches 462; Indels 37; Gaps 12;
QY	1	TCGCTGTGAGCTCGAATATGCTGCGGATCCGCGCGCGGAGCTGTTAAGAG 60
Db	35	TCGCGGGGTTTCGGGAAATGCTGCGAGACCCCTAGAGGCTCTGCGAGCTTACCTCAGG 94
QY	61	CAGTCCCACTGTCTCCCTTGTGTTTTCGAGAGCTCCGAAATCTTCTCATGTATGAACTGTG 120
Db	95	GAAACAGCTCTAGTATATCTAGATGTTGAAATATC-GAAGCTGTACTGTGAAACAGT 153
QY	121	ACTGACAACTGAGGCT--GAGGCTTGCGCTCTGCTTG-----ACCTGAGCCCC 167
Db	154	GCTGACAAAGATGTTGTGTATGAGCTGCTGCTGTGCTTGGAACCAAGATTTCTGCTGCC 213
QY	168	AGGCTATATATTTATATGAGATTTCCAGCCAAAGTTTCCAACTTCACTTTTGGAGC 227
Db	214	AGGCTTTTGTGTATTTATTTAGATTT-CAGGGAAGATGTTCCAACTTTCAG-TGTGAGGC 271

```

QY 228 AGGTATGATGATTAAGGCACTGAGCAATGAGAAAGCAACCAAGCTATCAAAACCAC 287
DB 272 AGGTATGAGAGCAAAAGCCGAGCGAATGAGAGCACTTAAGGCAATCAAAACCAC 331
QY 288 CAGTAAGAGATTCAGAAAACCTGGGGTTTTCAGAAACCAAGTTGCCAAACGTGAGG 347
DB 332 CAGCAAAAGATTCAGAAAACATGGGGTTTTCAGAAACCAAGTTGCCAAACGTGAGG 391
QY 348 TGCAGAGACGAGAGCGAGCCCACTGAGCAACA-----ACCAAGACAGCAATTAACCTC 402
DB 392 CGCAGGGAGCGCGAGGCTGACCCCACTGAGCCCAACCCCAAGCAGCAGCACTGGGCTG 451
QY 403 TCCCTGCGCGCACTGAGACGCAACCAAAAGTACTGAGAGGTTGAAGATTTCTTAAC 462
DB 452 TCCCTGCGCGCACTGAGAGCGCAACCAAGCACTGAGAGCGAGTGAAGCAATTTCTGAC 511
QY 463 ACAGTTGCGCGCCGAGGGAGAAAAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCC 522
DB 512 ATTGCGCGCGCCGCGAGAGAGAGCAATGCTGCTCTCTGAGAGATTTCTGAGGCCC 571
QY 523 ACATCTTCACAGTCACTGATGTGAGACAGCTTCCAGGGAGCGTTGAAGCAATTTCT 582
DB 572 ACCTCTGCGCGC--CAACAGCGCCGAGACAGC--TCCGAGGCGAGGTGAAAAGCGCTTCT 629
QY 583 GAGATCAAGATGCGCCCTGTATCTGACTCCCTTAAGGAAAGAACTCCGCGCTTTCTGAA 642
DB 630 GAGATCAAGAGCGGCGCCCAAGTCTCTTCAAGTGTGAAGAGCAACAGCTCTTCTGA 689
QY 643 AAGCAAAAGAGGTGAAAGAGAAAGACACTCTGACAGTGAAGTATGAGCTTAAG 702
DB 690 AAGGTGAAGAGAGGAGATGACCAAGATGACCTCCGATGATGACAGGATGAG--CTGACC 748
QY 703 TTGAAGAACTTCAGAACCGCTTTCGAGAAAGCGAGCAAGAACTGTGAGAGGTCC 762
DB 749 TTGAAGAGCTTCAGAAATCG--CTTCGACAGAAAGCGGAAACAGAGCCCACTGAGAGGCCC 807
QY 763 CTGAGAGGCACTCAAGATTCGCTGAGAGAAAGCGAGAGAAAGATTTGCCGAACCT 822
DB 808 CTGAGAGGATTCAGAGATTCGCTGAGAGAAAGCGCGAGAGAGGTCCGCGCAAGCT 867
QY 823 GGGAGTGTCCAAATAGGCAAGTCCGAGAGAGACAGACCTCTCTGTAGCAGAGGCTTGA 882
DB 868 GTGGGCTCCGAGGCGCAAGTGAACATGTGAGAGGCGTCTGCGCAAGTAAACAGAGCCGAG 927
QY 883 CCTAGTCAGGAGCAAGTGTCCCACTCAGAGACAGTGAATGAATGAATCAAGTTGAGAGG 942
DB 928 AACGATCAGGGGGTGTGTCTCCAGGCTGGAAGATGACAGAGAGTGAATTTGAGAGGA 987
QY 943 AAGCGCACTCAGGGAAATATCAGAGAGAAACCAGGAAAGCGGCAAAACCAAGGCTGAG 1002
DB 988 AAGGCGGCTCAGAGACATCAAGATGAGAGGCTGAGAGCTTGGGCGAGACCGAGAGCTGAA 1047
QY 1003 TGTGAGGTTTACGACCCCAATGCTCTGTACTGCACTGCGCCGACGCTCACAACCAAG 1062
DB 1048 TGTGAGGTTTACGACCCCAAGGCTCTGTATTGCAATTTGCCGACGCTCACAACCAAG 1107
QY 1063 TTTATGATCTGCTGTGATCGGTGTGAGAGAGTGTTCATGTGATCTGTGTGGTATTTCT 1122
DB 1108 TTTATGATTTCTGTGACCGGTGTGAGAGATGTTTCAATGTGAGATTTGTGGGCAATTTCT 1167
QY 1123 GAGGCGGAGGGCGGCTCTGGAAGAAAGAAACCGGAAAGACTAATCTGCGCCAAATTTGCAAC 1182
DB 1168 GAGGCTCAGAGGAGGCTTTTGAAGAAAGATGGGAAAGACTAATCTGCGCCAAATTTGCAAC 1227
QY 1183 ATTTTGAAGTGCAGAGATGAGCAAAACGCTAGCGCCCAACATGAGCAGAGACTTGGGTGC 1242
DB 1228 ATTTGCAAGTGCAGAGATGAGCAATTCAGAAACGCGCAGATCAGCAGAGAACTTAATGG 1287
QY 1243 AGATCTGTGGGTGTGATGAGCAACAGCTGCAACAGATGAGGACAGTGAAGCAAGTCC 1302
DB 1288 AGACTGTGAGATGCTGATGAGCAACGATTTGTACAAATATATGAGAAATATGAGCAAGTCT 1347

```

```

QY 1303 GAGAGACGAGGCAATTAAGGATGAGAGGCAAGCAACCCAGCGGCAAGAA 1362
DB 1348 AGCAGAGACCAAGGATTAAGGATGAGAGGATGAGAGGATGAGAGGATGAGAGG 1407
QY 1363 AAACCTCAAGATTAATCAAGCTGTCTGAGAGGCTCTGAGTCTCTTAATGATTTGGCCT 1422
DB 1408 AAACCTCAAGATCTTCCAGGCTGTGATGAGAGGCTCTGAGTCTCTTAATGATTTGGCCT 1467
QY 1423 GGGTGTCCAGTGAAGCAAGCGCTGAGTGTGATGATGATGATGATGATGATGATGAT 1482
DB 1468 GGGTGTCTCAGTGTGAGAGGCGGAGCTCGGTGTGATGATGATGATGATGATGATGAT 1527
QY 1483 CAGCAGCAGCTACATGATGATTTCTAAGTTCAAGTAAAGAAACAAACAAACCAAG 1542
DB 1528 CAGCAGCAGGAGCAATGAGATTTCTAAGTTCAAGTAAAGAAACAAACCAAGCTTAA 1587
QY 1543 GAAAAGCTCAAGAGAGAAAGCCAGAAAGTTCAAGTCTTCAAAATGCAATGATGATGAG 1602
DB 1588 GAAAAGATGAGATGAGAGCCAGAGAGCCAGTCTTCCGAAATGCGGTGCTCAGGAGGT 1647
QY 1603 ATTAATATCTCTCTGAGCAAGAGACTAGCGTCAAGAGAAAGGAGAAACCCAGTGAAG 1662
DB 1648 ATTAATATCTCTCTGAGCAAGAGAGCTAGCGTCAAGAGAAAGGAGAAACCCAGTGAAG 1707
QY 1663 AA--AGTATGCTGCTTCCAGAGATGAGACTTCTGAGAAAGAGGCAAGCTGTGAGAGC 1719
DB 1708 AAGCAGTGTGTGTGCTTCCAGAGATGAGCACTTCCGAAAGAGAGAGCTGTGAGAGC 1767
QY 1720 AGCAGCAGCTCTGAGGAGAGTGAACCAATCAATGCTGTGAGAGGAGAGAGGCTGAG 1779
DB 1768 AGCAGCAGCTCTGAGGAGAGTGAACCAATCAATGAGTGAAGAGGAGAGAGGCTGAG 1825
QY 1780 AAGCCCACTGCACTCTGCGCCCAACCTTATGAGTAAATGATGATGATGATGATGATGAT 1839
DB 1826 -----CTGCTCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1878
QY 1840 TTCCAGGCGCCCTCCATCATCT 1862
DB 1879 CTCTGAGACCTCTCCGTTCTTT 1901

RESULT 6
ID ABV22095
XX ABV22095 standard; cDNA; 2085 BP.
XX
AC ABV22095;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 22086.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189863P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX

```


QY 1780 AAGCCACATGACTCTGCGCCACCCCTATTAGTAATGTAGTATACACCAAGGCTGGC 1839
DB 1826 -----CTGCTCCTCGCCCTGACTGTGTAAATATATATACCTAGGGGTGGC 1878
QY 1840 TCCCAAGGCCCCCTCCATCATCT 1862
DB 1879 CTCCTGACCCCTCCGCTTCTTT 1901

RESULT 7
ABV22192
ID ABV22192 standard; cDNA, 2085 BP.
AC ABV22192;
XX
XX 13-SEP-2002 (first entry)
DT
DE Human prostate expression marker cDNA 22183.
XX
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX Homo sapiens.
OS
XX WO200160860-A2.
PN
XX 23-AUG-2001.
PD
XX 20-FEB-2001; 2001MO-US005171.
PF
XX
XX 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0239007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 3816; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
CC
XX
SQ Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;

Query Match 33.1%; Score 948.2; DB 5; Length 2085;
Best Local Similarity 73.4%; Pred. No. 1.3e-285;
Matches 1383; Conservative 0; Mismatches 463; Indels 37; Gaps 12;

QY 1 TCCGTGAGTCTTGAATGCTGCGGATCCCGGCGCGGAGCTTGTTTAAGAG 60
DB 35 TCCGCGGCTTCGGGGAATGCTGCGAGACCTGAGGCGCTGCGAGCTTACTCCAGG 94

QY 61 CAGTCCCACTGTCCTTGTGTTGTCGAAGCTCCGGAATCTTCTCATGTATGAATCTG 120
DB 95 GAACAGCCTTGAATATCTGATGTGTAAATATC-GAAGCTGTATCTCGGAACAGTG 153
QY 121 ACTGACAAAGTGGGT--GAGCTTGCCCTCTCTTG-----ACCTGCCCC 167
DB 154 GCTGACAAAGTGTGTGTGTGAGCTCGCTGTCTCTGACCCAGAGGTTCGTGCTG 213
QY 168 AGGTCTAATATTTATGTATGATTTCCAGCCCAAGGTTTCCAGGTTTCAGTTTGGAC 227
DB 214 AAGGTTTGTGTATTTATTAAGATTT-CAGGAAAAGTGTCCAGCTTTCAG-TGTGAGC 271
QY 228 AGGTATGATGATTAAGGGAACCTGAGCAATGAGGAACCAAGGCTATCAAAACCCAC 287
DB 272 AGGTATGACACAAAGGCGACCCGACAAATGAGAGGACCTTAAGCATCAAAACCCAC 331
QY 288 CAGTAAGAGTTTCAGAAAACCTGCGGTTTTTCGAAGAACCAAGATTGCCAAACGTGAGG 347
DB 332 CAGCAAAAGATTCAAGAAAACATGCGGTTTTTCGAAGACCACTATGCGCAACGAGAGGG 391
QY 348 TGCAGGAACACAGAGGCGGACCCCAAGTGAAGCA-----ACCAAGAGCATTAACCTC 402
DB 392 CGCAGGAGACGCGGAGGCTGACCACTGAGCCGACCCCAAGCAGCAGCTGCGGCTG 451
QY 403 TCCCTGCGCCGAGTGAACGCGCAACCAAAACGTAAGAGGAGTGAAGATTCTTAC 462
DB 452 TCCCTGCGCCGAGTGAACGCGCAACCAAAACGTAAGAGGAGTGAAGATTCTTAC 511
QY 463 ACGGTTTCGCGCCGAGGAGAAAAGATGTGCGGTTCTCTGAGAGATTCAAGTGACCC 522
DB 512 ATTCGCGCGCGCGGAGGAGAGCACTGCTCTCTCTGAGAGATTCTGAGGCCC 571
QY 523 ACATTTTCAAGTACATGATGTGAGACAGCTTCCGAGGAGGAGGTTGAAGCACTTCT 582
DB 572 ACGTCTGCGCCG-CACAGAGCGCTGAGCAG-TCTGAGGAGGAGGTTGAAGCGCTTCT 629
QY 583 GAGATCAGAAAGTGGCCCTGTATCTGACTCTTGAAGAAAGAACATCTGCTCTTCTGAA 642
DB 630 GAGACCAAGAGCGGCCCAAGTCTGCTTCCACAGTGTGAAGAGACAGACAGCTCTTCTGAA 689
QY 643 AAGCAAAAGAGGTGAAGAGAGAGAGAACACTCTGACAGTGAAGTGAAGGCTTACG 702
DB 690 AAGGTGAAGAGGAGGAGTGAACCAAGTACACCTCCGATGAGCAGGATGAG-CTGACC 748
QY 703 TTGAAGGAATCTTCAAGACCGCTTCTGAGAAAGGCAAGAGCAAGACCTTGTGAAGAGTCC 762
DB 749 TTGAAGAGCTTCAAGATCG-CTTCGACAGAAAGCGGAAACAGAGCCCACTGAAGGCCC 807
QY 763 CTGAGAGGAGTCAAGATCGCTGAGAAAGGCGCAGAGAGAGATTTCTGCCGAATCT 822
DB 808 CTGAAGAGGATTCAGAGTGTGCTGAGAAAGAGCCCGGAGAGAGGTTCCCGCCAGACT 867
QY 823 GGGAGTGTCCAAATAGGAGATGTCGAGAGAGACAGACCTTCTGTGAAGAGAGCTGAG 882
DB 868 GTGGCTCGAGAGCGCAAGTGAACCTGTGAGGCGGTCTGCCAGTGAAGAGAGCCGAG 927
QY 883 GCTAGTCAAGGACCAAGTGTCCAGTCAAGACAGATGACATGTGAAGATGATGTTGAAGG 942
DB 928 AACGATCAGGGGGTGTGTCCAGGCTGGAAGATGACAGAGAGATGAAGTTGAAGGGA 987
QY 943 AAGGGAATCAAGGAAATACAGAGAAAACCCAGGAGAGCGGGCAACCAAGCTGAG 1002
DB 988 AAGGCGCTCAGAGACATCAAGATGAGAGCTTGAAGACTTGGCGAGCCGAAGCTGAA 1047
QY 1003 TGTGAGTTTACGACCCCAATGCTGTACTGATCTGCGGCGAGCTTCAACAACAG 1062
DB 1048 TGTGAGGTTTACGACCCCAATGCTGTACTGATCTGCGGCGAGCTTCAACAACAG 1107
QY 1063 TTTATGATCTGCTGTGATGCTGTGAGAGAGTGTTCATGTGATCTGTGTGTTTCT 1122
DB 1108 TTTATGATTTGTGTGACCGCTGTGAAAGATGTGTTTCAATGAGATGTGTGAGCAATTTCT 1167
QY 1123 GAGGCGCAGAGGCGGCTCTGTGAAGAAAGAGAGGAGAAAGATCAATCTGCTCAATTTGACC 1182


```
QY 833 AAATAGCAGTGGCGAGGAGCAGACCTCTCTGTAGCAGAGGCTGAGCTAGTCAG 892
DB 484 AGCCGAGTGAACCTGTGAGGAGGCGCTCTGCTCAGTACAGAGGCCAGAAACATCAG 543
QY 893 GACCACTGTCCAGTCAAGACAGATGACATAGAAATCACTGGAGAGGAGGACTC 952
DB 544 GCGTTGTGTCCAGGCTGGAAAGATGACAGAGATGATGAGTGGAGGAAAGGCGCTC 603
QY 953 AGGAAATACAGAGGAAACCCAGGAGAGGCGGAAACCAAGCTGATGATGAGTT 1012
DB 604 AGGACATCAAGATGAGAGGCTGAGAGCTTGGGCGAGCCGAGGCTTAATGTAGGGTT 663
QY 1013 AGCAGCCCAATGCTCTGATGACATCTGCGCGAGCTCAGACAAAGGTTTATGATCT 1072
DB 664 AGCAGCCCAAGGCTCTGATGATGATCTGCGCGAGCTCAGACAAAGGTTTATGATCT 723
QY 1073 GCTGTGATCGGTGTGAGAGAGTGTTCATGATGATGATGATGATGATGATGATGAT 1132
DB 724 GCTGTGATCGGTGTGAGAGAGTGTTCATGATGATGATGATGATGATGATGATGATGAT 783
QY 1133 GCGCGCTCTGAGAAAGGAGAGGAGAGAGTACATCTGCGCAATGACATCTTTCAG 1192
DB 784 GAGGCTTTTGGAAAGATGAGAGGAGAGTATATCTGCGCAATGACATCTTTCAG 843
QY 1193 TGCAAGATGAGCAAAAGGAGAGGAGAGTACATCTGCGCAAGATCTGAGTCTGAG 1252
DB 844 TGCAAGATGAGCAATCTATTCAGAAAGGAGAGTACATCTGAGAGAGCTTATGAGAGCTT 903
QY 1253 GTGCTGATGAGCAGACATGACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1312
DB 904 ATGCTGATGAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 963
QY 1313 AGGCAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1372
DB 964 AAGGATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023
QY 1373 TATTCAGGCT----- 1383
DB 1024 TCTTCAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1083
QY 1384 -----G 1384
DB 1084 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
QY 1385 TCGTATGAGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1444
DB 1444 TGAATGAGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1203
QY 1445 CTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1504
DB 1204 CCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1263
QY 1505 TTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1564
DB 1264 TTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1323
QY 1565 AAAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1624
DB 1324 AGAAGCCAGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1383
QY 1625 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1681
DB 1384 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1443
QY 1682 GAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1741
DB 1444 GAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1503
QY 1742 AGCAGCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1801
DB 1504 ATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1554
QY 1802 CCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1823
```

```
DB 1555 CACTGTGTATTAATCCAGAA 1576
RESULT 9
AAK52368
ID AAK52368 standard; cDNA; 7838 BP.
XX
AC AAK52368;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 913.
XX
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukemia;
KM nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PE 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663551.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HVS8-) HVS8Q INC.
XX
PI Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
XX
PT P-PSDB; AAM79235.
XX
PS Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.
XX
PS Claim 1, Page 3007-3013; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 7838 BP; 1672 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;
Query Match 25.1%; Score 719.2; DB 4; Length 7838;
Best Local Similarity 70.7%; Pred. No. 2,1e-213;
Matches 1076; Conservative 0; Mismatches 323; Indels 123; Gaps 4;
QY 416 GTGAGCGCAACCAAGCTGATGAGAGGATGATGATGATGATGATGATGATGATGATGAT 475
DB 599 GTGAGAGCAACCAAGCTGATGAGAGGATGATGATGATGATGATGATGATGATGATGAT 658
```


QY 476 GAGGAGAGAGAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCACATCTTCCACAG 535
 DB 659 GCGGAGAGAGAGAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCACATCTTCCACAG 718
 QY 536 TCACAGATGAGAGAGATTCGAGAGAGAGAGATTCAGTGAAGCCACATCTTCCACAG 595
 DB 719 CCACAGAGAGAGAGATTCGAGAGAGAGAGATTCAGTGAAGCCACATCTTCCACAG 778
 QY 596 GCCCTGATGATGATCTTCA---GGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 652
 DB 779 GCCCTGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 838
 QY 653 GAGGTGAAGAGAGAGATTCGAGAGAGAGAGATTCAGTGAAGCCACATCTTCCACAG 712
 DB 839 GAGGTGAAGAGAGATTCGAGAGAGAGAGATTCAGTGAAGCCACATCTTCCACAG 898
 QY 713 TTCAAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 772
 DB 899 TTCAAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 958
 QY 773 GTCAAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 832
 DB 959 TCAGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1018
 QY 833 AAATAGAGAGAGATTCGAGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 892
 DB 1019 AGGCGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1078
 QY 893 GACGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 952
 DB 1079 GAGGTGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1138
 QY 953 AGGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1012
 DB 1139 AGGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1198
 QY 1013 AGGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1072
 DB 1199 AGGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1258
 QY 1073 GCTGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1132
 DB 1259 GCTGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1318
 QY 1133 GAGGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1192
 DB 1319 GAGGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1378
 QY 1193 TGACAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1252
 DB 1379 TGACAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1438
 QY 1253 GTCGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1312
 DB 1439 GTCGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1498
 QY 1313 AGGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1372
 DB 1499 AGGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1558
 QY 1373 TATTCAGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1383
 DB 1559 TATTCAGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1618
 QY 1384 TATTCAGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1384
 DB 1619 AAATGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1678
 QY 1385 TCGATGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1444
 DB 1679 TCGATGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1738

QY 1445 CTGACTGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1504
 DB 1739 CTGACTGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1798
 QY 1505 TTCTAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1564
 DB 1799 TTCTAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1858
 QY 1565 AAAAGTGAAGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1624
 DB 1859 AAAAGTGAAGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1918
 QY 1625 AGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1681
 DB 1919 AGAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1978
 QY 1682 GAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1741
 DB 1979 GAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 2038
 QY 1742 ACCAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1801
 DB 2039 ACCAGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 2089
 QY 1802 CCTATGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 1863
 DB 2090 CCTATGATGATGATCTTCA---GGAGAGAGAGATTCGAGAGAGAGATTCAGTGAAGCCACAT 2111
 RESULT 10
 AAS86576
 ID AAS86576 standard; cDNA; 2332 BP.
 AC AAS86576;
 AC AAS86576;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #22380.
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US008631.
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 PA (HYSB-) HYSBQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR MPI; 2001-639362/73.
 DR P-PSDB; ABG22389.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostic, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PS Claim 1; SEQ ID NO 22380; 103bp; English.
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostic as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

XX Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;
 SQ
 Query Match 24.1%; Score 692; DB 5; Length 2332;
 Best Local Similarity 71.3%; Pred. No. 3,2e-205;
 Matches 1051; Conservative 0; Mismatches 335; Indels 88; Gaps 7;

QY 211 GCTTCACTTTTGGAGCAGTATGATGATTAAGGCACTGAGCAATGAGAGACCC 270
 DB |||||
 DB 370 GCTTCACTTTTGGAGCAGTATGATGATTAAGGCACTGAGCAATGAGAGACCC 429
 QY 271 AAGCTTCAACCCCAACCAAGTAAAGTTCAGAGAAACCTGGGGTTTCCAGAAACCA 330
 DB |||||
 DB 430 AAGGCTCAACCCCAACCAAGTAAAGTTCAGAGAAACCTGGGGTTTCCAGAAACCA 489
 QY 331 ATTCCAAACGTAGAGGTGAGAGACAGGAGCGGACCCCACTGAGCAGCAACCC 386
 DB |||||
 DB 490 ATTCCAAACGTAGAGGTGAGAGACAGGAGCGGACCCCACTGAGCAGCAACCC 549
 QY 387 --ACAGCAGATTAACCTCTCCCTGCGCGCAGTGAACGCAACCAACCTAATGAGAG 444
 DB |||||
 DB 550 CAGCAGCAGCTGAGCTGTCCCTGCGCGCAGTGAACGCAACCAACCTAATGAGAG 609
 QY 445 GTAAAGAGTTTCTTAACACGCTTCCGCGCGGAGGAGAAAGTATCCGCTGCTCC 504
 DB |||||
 DB 610 GTGAAGAGTTTCTTAACACGCTTCCGCGCGGAGGAGAAAGTATCCGCTGCTCC 669
 QY 505 GAGATTCAGTGAAGCCCACTCTTCAAGTCACTGATGATGAGAGCAGTTCAGAGAG 564
 DB |||||
 DB 670 GAGATTCAGTGAAGCCCACTCTTCAAGTCACTGATGATGAGAGCAGTTCAGAGAG 729
 QY 565 AGCTTGAAGCAAGTTCTGAAGTCAAGAGTGGCCCTGATCTGACTCTTA---GGAAA 621
 DB |||||
 DB 730 AGCTTGAAGCAAGTTCTGAAGTCAAGAGTGGCCCTGATCTGACTCTTA---GGAAA 789
 QY 622 GAACATCTGCTCTTCTGAAGAGCAAGAGGAGTGAAGAGAGAGCACTCTGAC 681
 DB |||||
 DB 790 GAACATCTGCTCTTCTGAAGAGGAGTGAAGAGGAGTGAAGAGAGAGCACTCTGAC 849
 QY 682 AGTGAAGTGAAGGCTTACCTTGAAGAGCACTGAGAAACCGCTTCGAGAAAGCAGAG 741
 DB |||||
 DB 850 AGTGAAGTGAAGGCTTACCTTGAAGAGCACTGAGAAACCGCTTCGAGAAAGCAGAG 909
 QY 742 CAAGAACTGTGAGAGGCTTCCCTGAGAGCACTGAGAAACCGCTTCGAGAAAGCAGAG 801
 DB |||||
 DB 910 CAGAGAGCCCACTGAGAGGCTTCCCTGAGAGCACTGAGAAACCGCTTCGAGAAAGCAGAG 969
 QY 802 GAGGAAGTTTGTCCGGAACCTGGAGAGTCCAAAGAGCAGTCCGAGCAGAGCAGAGACT 861
 DB |||||
 DB 970 GAGGAAGTTTGTCCGGAACCTGGAGAGTCCAAAGAGCAGTCCGAGCAGAGCAGAGACT 1029
 QY 862 CTCTGTAAGCAGAGCCTTGAAGCTTGAAGCACTGAGTCCAGTCAAGAGCAGATGAC 921
 DB |||||
 DB 1030 CCAAGTAAGCAGAGCCTTGAAGCTTGAAGCACTGAGTCCAGTCAAGAGCAGATGAC 1089
 QY 922 ATAGAAATCACTGTAAGAGGAGGAGCAGTCAAGGAAATACAGAGAGAAACCCAGGGA- 980
 DB |||||
 DB 1090 AGAGAGTAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1149

QY 981 AGCGGCAAAACCAAGCTGAGTGAAGTTTACACCCCAATGCTGTACTGATCTG 1040
 DB |||||
 DB 1150 TTGGGGCCAGACCGAAGCTGAAATGAGAGGTTTACACCCCAATGCTGTACTGATCTG 1209
 QY 1041 CCGCAGCTTCAACAACAACAGTTTATGATCTGCTGTAATGAGTGAAGAGTGTTC 1100
 DB |||||
 DB 1210 CCGCAGCTTCAACAACAACAGTTTATGATCTGCTGTAATGAGTGAAGAGTGTTC 1269
 QY 1101 TGGTACCTGTGTGATTTCTGAGGCGCGAGGCGCTCCGAGAAAGAAAGGAGAG 1160
 DB |||||
 DB 1270 TGGCATTGTGTGATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGAAATGAGAG 1329
 QY 1161 CTACATCTGCCAAATTTGACCAATTTTCAAGTGAAGTGAAGCAAAACGTTAGCGCCAC 1220
 DB |||||
 DB 1330 CTATATCTGCCAAATTTGACCAATTTCTGCAAGTGAAGTGAAGTCAATTCAGAAACGCG 1389
 QY 1221 CAATGAGCAGAACTCTGGGTCAAGTCTGTGTGCTGATGAGCAGACAC---TGACCA 1275
 DB |||||
 DB 1390 AGATCAGCAGAAAGCTTAAATGAGAGCTGTGAGATCTTGAATGAGCAGATTTTAC 1449
 QY 1276 AGCATAGGAGACGT--AGAGCAAGATC---CGAGAAAGCAGAGGATTAAGGGTACG 1329
 DB |||||
 DB 1450 AGTATAGGAAACATTAGAGAGAGAGTCTTACCGAGAACCCAGGGGTTAAAGGTTAC 1509
 QY 1330 ATTGAGAGCAGCAAAACCCAGCGCAAGAAAGAAATCAAGATATTCAGAGC----- 1382
 DB |||||
 DB 1510 ATTGAGAAAGCTGCAAAATCAAGTGGCAAGAAAGAAATCAAGATTTTCCAGCTGTAT 1569
 QY 1383 ----- 1382
 DB 1570 GAAATTCGTGTCTGATGAGACATCTGCTTCACACTTGCACATGATAGCTGCA 1629
 QY 1383 TGTGTGAGAGCTCTGTGTGCTCTTAATGATAGGAGCTGTGCTCAAGTATGACCA 1442
 DB |||||
 DB 1630 GGTATAGAGAGCTGTGTGCTCTTAATGATAGGAGCTGTGCTCAAGTATGACCA 1689
 QY 1443 GCTGACTCTGTGTATGTCAGTAAATGATGATCTCAAAACAGCAGCAGTCAATGAG 1502
 DB |||||
 DB 1690 GCCGACCTCGGTGTATGCAATGATGATGATCTCAAAACAGCAGCAGCAGTCAATGAG 1749
 QY 1503 ATTCTAAGTTCAGGTAAGAAACAAACAAACCAAGGAGAAAGTCAAGACAGAGC 1562
 DB |||||
 DB 1750 GTTCTAAGTTCAGGTAAGAAACAAACCAAGGAGTCAAGGAGTCAAGATGAGAGC 1809
 QY 1563 AGAAAGTTCACTCTTCCAAATGCAAGTTCAG 1596
 DB |||||
 DB 1810 AGAGAAAGCCAGTCTTCCGAAATGCGGTCTCAG 1843

RESULT 11
 AAC77933
 ID AAC77933 standard; cDNA, 1764 BP.
 XX
 AC AAC77933;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated gene sequence SEQ ID NO:327.
 XX
 XX Human: cancer associated gene; cancer antigen; detection; cancer;
 KM diagnosis; cytotoxic; proliferative; vulnerability; immunomodulator;
 KM antidiabetic; antiaesthetic; antithrombotic; antidiabetic; antidiabetic;
 KM antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;
 KM dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KM vasorelaxant; antiproliferative; angiogenic; gene therapy; inflammation;
 KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KM allergic reaction; graft versus host disease; organ rejection;
 KM haemostatic; thrombolytic; cardiovascular disorder; infection;
 KM neurological disease; drug screening; ss.
 XX
 OS Homo sapiens.
 XX

WO200055550-A1.
 21-SEP-2000.
 08-MAR-2000; 2000WO-US005882.
 12-MAR-1999; 99US-0124270P.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Ruben SM;
 WPI; 2000-587533/55.
 P-PSDB; AAB43724.
 Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer.
 Claim 1; Page 880-881; 2352pp; English.
 AAC7607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44233. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulnery; immunomodulator; antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral; dermatological; neuroprotective; cardiant; thrombolytic; coagulant; neurotropic; vasotropic; antipsoriatic and antiangiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells; to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC7849 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention

Seq	Sequence	1764 BP; 468 A; 424 C; 415 G; 453 T; 0 U; 4 Other;
Qy	Query Match	19.5%; Score 558; DB 3; Length 1764;
Db	Best Local Similarity	74.2%; Pred. No. 2,8e-163;
	Matches 795; Conservative 0; Indels 27; Gaps 6;	
Qy	984 GGGGAAACCAAAGCTGAGTGTGAGGTTTACGACCCCAATGCGCTGTATCTGCATCTGCGG	10433
Db	32 GGGGCGACCGAAGCTTAATGTGAGGGTTAGACCCCAACCGCTGTATCTGCATTTGCGG	91
Qy	1044 CCAAGCTCACAAACAAGTTTATGATCTGCTGTATCGGTGTGAGAGTGGTTCATAG	11030
Db	92 CCAAGCTCACAAACAAGTTTATGATTTGCTGTGACCGCTGTGAAGATGTTTCAATAG	151
Qy	1104 TGACTGTGTGGGTATTTTCTGAGGCGCCGAGGGCGGCTCTGTGAAGAGAAACGGGAAAGCTA	1163
Db	152 CGATTGTGTGGGCTATTTCTGAGGCTCGAGGGAGGCTTTTGTGAAGAAATGGGAAAGCTA	211
Qy	1164 CATCTGCCCAATATGCAACATTTTGTCAAGTGCAGATATGAGCAAAACGGTAGCGCACCA	1223
Db	212 TATCTGCCCAATCTGCACCAATTTCTGCAGATGCAGATATGAGCTCATTTCAAGAAACGGCAGA	271
Qy	1224 TGAACGAGACTTGTGGGTGCAGATGTGTGGGTGTGATATGACACAGATGCACACATATAG	1283
Db	272 TCAGCAGAGAGCTTAATATGAGAGCTGTGAGATGTCTGATATGCAACCATTTGTACAAGTATATAG	331
Qy	1284 GACAGTGAACAGAAAGTCCCGAGAAAGACAGGGCATTAAGGGTATGAGATTGAGAAAGCGAG	1343
Db	332 AACATATGAGAGAAAGTCTTACCGAAGACCAAGGGATTAAGGGTATGAAATTGAGAAAGCTGC	391
Qy	1344 AAACCCGAGGCAAGAAAAAATCAAGATATTTCAAGCTGTGTGTAGAGGCTCTGTGGTC	1403

[illegible]

XX	RESULT 12
XX	AAA43780
XX	ID AAA43780 standard; cDNA; 408 BP.
XX	XX
XX	AC AAA43780;
XX	DT 21-AUG-2000 (first entry)
XX	XX
DE	Mouse secreted expressed sequence tag SEQ ID NO:355.
XX	XX
KM	Human; mouse; chicken; rat; secreted expressed sequence tag; EST;
KM	expressed sequence tag; EST; probe; chemotactic; proliferative; SEST;
KM	immunomodulatory; haematopoietic; chemokine; anagastic; haemostatic;
KM	thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KM	antiviral; antidiabetic; antilastmatic; vulnerary; antiparkinsonian;
KM	antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
KM	cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
KM	autoimmune disorder; multiple sclerosis; allergic condition;
KM	insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KM	lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KM	central nervous system disorder; Alzheimer's disease; stroke;
KM	Parkinson's disease; Huntington's disease; coagulation disorder;
KM	haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
KM	infection; depression; psoriasis; ss.
XX	XX

QY 924 AGAATAATCGATTGAGAGGAGGCGACTGACGGAAATACAGAGAAAACCCGAGGAGC 983
DB 2 AGAGAGTAGTGGAGGAGGAGGCGCTCAGGACATCAAGATGAGAGGCTTGAGACTT 61
QY 984 GGGCAAAACCAAGCCCTGAGTGTGATTTACACCCCAATGCCCTGTACTGTGATCTGCGC 1043
DB 62 GGGCCGACCGAAGCCCTGAATGTGAGGTTTACGACCCCAACGCGCTGTATTTGCAATTTGCCG 121
QY 1044 CCAGCTTCACAAACAAGGTTTATGATCTGCTGTGATCGGTGTGAGAGTGTTCATG 1103
DB 122 CAGGCTTCACAAACAAGGTTTATGATTTGCTGTGACCGCTGTGAGAGATGTTTCAATG 181
QY 1104 TGACTGTGTGATTTTGAAGGCGGAGCGGCTCTTGAGAAAGAAACGGGAGAGACTA 1163
DB 182 CGATTGTGTGGCATTTCTGAGGCTCGAGAGGCTTTTGAGAAAGAAATGGGAGAGACTA 241
QY 1164 CATCTGCCCAATTTGACATTTTGAAGTGTGAGAGTGTGAGCAACGGTACGCCCAACAA 1223
DB 242 TATCTGCCCAATTTGACATTTCTGAGGCTCGAGAGGCTTTTGAGAAAGAAATGGGAGAG 301
QY 1224 TGAGAGAGACTCTGGGTGAGATCTGTGGGTGCTGATGGCAGACTGCAACAGCATAG 1283
DB 302 TGAGAGAGAGACTTAAATGAGAGCTTGAGATGCTGATGGCAGACTTGTACAAATATAG 361
QY 1284 GACAGTAGAGAGAGAGAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1343
DB 362 AACAAATAGAGAGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
QY 1344 AAACCCGAGGCGAG 1383
DB 422 AAATTCAGAGTGTGAG 481
QY 1384 ----- 1383
DB 482 CCAAGCTGCTGTCTCTGCGAGAGTATGAGAAATGCTGTGTCTAGAGAGAGATCTGCTT 541
QY 1384 -----GTCTGAGAGAGCTCTCTGTGCTCTTAAATGAT 1415
DB 542 CACAGCTTGTGACCTGATTAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
QY 1416 TGG-CCCTGGGTGTCTCAAGTGTGAGCAG-CTGTGATCTGTGTATTTGCA-GTATGACTG 1472
DB 602 TGGCCCGGGGTGTCTCAAGTGTGAGCAG-CTGTGATCTGTGTATTTGCA-GTATGACTG 661
QY 1473 CATTTCTAAACGAGAG-AGCTACATGAGATTTCTAAGTTGAGGTAAAGAAACAAA 1531
DB 662 TATCTCTAAACGAG 721
QY 1532 CAAAACCCAG 1591
DB 722 CAAAGCTTAAAG 781
QY 1592 TTGAGGTGGAGTTAAATCTTTCTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1651
DB 782 CTCAGGAGAGTTAAATCTTTCTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
QY 1652 ACCGAGTGAAGAA---AGTATGCTGTGCTTCCAGAGTGAAGCTTTGAGAGAGAGAGAG 1708
DB 842 CCACAGTGAAG 901
QY 1709 CCTGTGAG 1768
DB 902 CTTGTGAG 955
QY 1769 AGAAGCCAG 1800
DB 956 TAAAGCCAG 987

RESULT 14
AAS86574 standard; cDNA; 503 BP.
ID AAS86574

XX AC AAS86574;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #22378.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN M0200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001MO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Dmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG22387.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostic, forensic, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1, SEQ ID NO 22378; 103bp; English.
XX XS

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantifying a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX XX

SO Sequence 503 BP; 165 A; 107 C; 138 G; 93 T; 0 U; 0 Other;

Query Match 12.1%; Score 347.8; DB 5; Length 503;
Best Local Similarity 80.7%; Pred. No. 1e-97;
Matches 406; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1132 GGGCGGCTCTGTGAG 1191
DB 1 GGGAGGCTTTTGAAG 60
QY 1192 GTGCAAGATGAG 1251
DB 61 GTGCAAGATGAG 120
QY 1252 GGTGCTGATGAG 1311

Db 121 GATGCTATGACACCGATTGTACAGATATAGGACATATAGACAGAGTCTAGCGAGAC 180
 QY 1312 CAGGGCATTAAGGGATGAGATTGAGAGGACGAAACCCAGCGGGAAGAAAACCTCAAG 1371
 Db 181 CAAGGATTAAGGGATGAGATTGAGAGGACGAAACCCAGCGGGAAGAAAACCTCAAG 240
 QY 1312 ATATTCCAGCCGCTGTAGAGGCTCTGGTCTCTTAATGATTTGGCCCTGGTCTCC 1431
 Db 241 ATCTTCAGCCCTGTATAGAGCGCTGGTCTCTTAATGATTTGGCCCTGGTCTCT 300
 QY 1432 AGTTAGACACGCTGACTCTGTATGACAGTAAATGATGATCTCTCAACACCGACGA 1491
 Db 301 CACGTGGCGGACCGGACCTCGGTACTGACAGTAAATGATGATCTCTCAACACCGGCA 360
 QY 1492 GCTACATGATGATTTCTTAAGTTCAAGTAAAGAAACAAAACCCAGAGAAAAGTC 1551
 Db 361 GCGAATAAGTATTTCTTAAGTTCAAGTAAAGAAACCCAGAGAAAAGT 420
 QY 1552 AAGACGAAGCCGAAAGGTTCAAGTCTTCCAAATGCAAGTGTTCAGGTGGGATTAATC 1611
 Db 421 AAGATGAAGCCGAGAGAGCCGAGTCTTCCAAATGCGGTGTCTCAGGCAAGTATTAAATC 480
 QY 1612 TCTTCTGTGACACAGAGACTAGC 1634
 Db 481 TCTTCTGTGACACAGAGACTAGC 503
 RESULT 15
 ID ABN21248 standard; cDNA, 324 BP.
 AC ABN21248;
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX polynucleotide sequence SEQ ID NO:10973.
 XX
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis; gene; ss.
 OS Homo sapiens.
 XX
 PN MO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US010836.
 XX
 PR 30-MAY-2000; 2000US-0206132P.
 XX
 PR 29-AUG-2000; 2000US-0228716P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinketsu RA, Leach MD;
 DR WPI: 2002-106308/14.
 DR P-PSDB: ABP05496.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders.
 XX
 ES Disclosure, SEQ ID NO 10973; 1037BP; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX

CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 324 BP; 106 A; 72 C; 78 G; 68 T; 0 U; 0 Other;

Query Match 11.2%; Score 322.4; DB 6; Length 324;
 Best Local Similarity 99.7%; Pred. No. 7e-90;
 Matches 323; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1300 TCCGAGAAAGCCAGGCGATTAAGGTTAGAGAAAGCGCAAAACCCAGCGGCAAG 1359
 Db 1 TCCGAGAAAGCCAGGCGATTAAGGTTAGAGAAAGCGCAAAACCCAGCGGCAAG 60
 QY 1360 AAAAACTCAAGATATTCAGGCTGTGTAAGAGGCTCTGTGCTCTTAATGATTC 1419
 Db 61 AAAAACTCAAGATATTCAGGCTGTGTAAGAGGCTCTGTGCTCTTAATGATTC 120
 QY 1420 CCTGGGTGTTCCAGTGTACAGAGCTGACTGTGTATGACAGTAATGATGATTC 1479
 Db 121 CCTGGGTGTTCCAGTGTACAGAGCTGACTGTGTATGACAGTAATGATGATTC 180
 QY 1480 AAAACGCGACGAGCTACATGAGATTTCTTAAGTCAAGTAAGAAACAAAACCC 1539
 Db 181 AAAACGCGACGAGCTACATGAGATTTCTTAAGTCAAGTAAGAAACAAAACCC 240
 QY 1540 AAGGAAAGGCTCAAGACGAAAGCGAAGAAAGTTCAAGCTTCCAAATGCAAGTGTTCAG 1599
 Db 241 AAGGAAAGGCTCAAGACGAAAGCGAAGAAAGTTCAAGCTTCCAAATGCAAGTGTTCAG 300
 QY 1600 GGGATTTAAATCTCTTCTGTGAC 1623
 Db 301 GGGATTTAAATCTCTTCTGTGAC 324

Search completed: April 23, 2004, 08:28:30
 Job time : 1074.35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:42:12 ; Search time 10916.3 Seconds
(without alignment) 11383.416 Million cell updates/sec

Title: US-09-787-016a-2

Perfect score: 2867
Sequence: 1 tccctgtagctcgtcgaat.....ttcttctctctgctgctg 2867

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: GenBank1:*
- 2: gb_ba:*
- 3: gb_hg:*
- 4: gb_in:*
- 5: gb_ov:*
- 6: gb_ov:*
- 7: gb_ov:*
- 8: gb_ov:*
- 9: gb_ov:*
- 10: gb_ov:*
- 11: gb_ov:*
- 12: gb_ov:*
- 13: gb_ov:*
- 14: gb_ov:*
- 15: gb_ov:*
- 16: gb_ov:*
- 17: gb_ov:*
- 18: gb_ov:*
- 19: gb_ov:*
- 20: gb_ov:*
- 21: gb_ov:*
- 22: gb_ov:*
- 23: gb_ov:*
- 24: gb_ov:*
- 25: gb_ov:*
- 26: gb_ov:*
- 27: gb_ov:*
- 28: gb_ov:*
- 29: gb_ov:*
- 30: gb_ov:*
- 31: gb_ov:*
- 32: gb_ov:*
- 33: gb_ov:*
- 34: gb_ov:*
- 35: gb_ov:*
- 36: gb_ov:*
- 37: gb_ov:*
- 38: gb_ov:*
- 39: gb_ov:*
- 40: gb_ov:*
- 41: gb_ov:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2867	100.0	2867	6	BD244713
2	2867	100.0	2867	6	AX023370
3	2602	90.8	2602	10	MMU28332
4	1770.4	61.8	4906	10	AK129117
5	1241.2	43.3	10463	10	AL732560
6	1023.6	35.7	2610	6	BD244712
7	1023.6	35.7	2610	6	AX023369
8	1015.2	35.4	2767	9	BC004237
9	1013.6	35.4	2725	9	BC014489
10	1013	35.3	2772	6	BD157094
11	1013	35.3	2772	6	BD157094
12	1013	35.3	2772	6	BD157094
13	985	34.4	2407	9	BC000770
14	853.4	29.8	221547	2	AC119716
15	748.2	26.1	272404	2	AC108337
16	703.4	24.5	69252	2	AC101519
17	449.8	15.7	6692	9	AB002331
18	365.4	12.7	28519	9	HSJ563814
19	324.4	11.3	69252	2	AC101519
20	230.4	8.0	160241	9	HS88517
21	230.2	8.0	2428	5	BC060442
22	209.8	7.3	681	6	AX86935
23	209.8	7.3	681	6	BD149997
24	184.6	6.4	812	11	BV013961
25	108.6	3.8	198133	2	BX296530
26	77.8	2.7	7218	6	I66494
27	70.2	2.4	61204	2	AC017132
28	70.2	2.4	161601	3	AC007594
29	70.2	2.4	225655	3	AB003695
30	69.6	2.4	2161	3	AK114522
31	57.8	2.0	2308	9	BC015733
32	57.8	2.0	2320	6	AX068654
33	57.8	2.0	2320	9	HS8801830
34	57.8	2.0	2331	9	BC014940
35	57.8	2.0	2387	10	BC030938
36	57.8	2.0	2452	9	AF149758
37	57.8	2.0	2487	9	AB031069
38	57.8	2.0	2509	10	AY099096
39	57.8	2.0	2967	9	HS8807333
40	56.8	2.0	1310	3	AY089443
41	56	2.0	10732	6	E32986
42	55.4	1.9	2888	9	AK055678
43	55.4	1.9	5975	6	AX329603
44	55.4	1.9	6256	6	AR338778
45	55.4	1.9	6936	9	D87685

ALIGNMENTS

RESULT 1	BD244713	2867 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD244713	Genes encoding for the human and murine death			
DEFINITION	BD244713	inducer-oblierator-1.			
ACCESSION	BD244713	GI:33054483			
VERSION	BD244713.1	GP 2002526040-A/2.			
KEYWORDS	JP 2002526040-A/2.				
SOURCE	Mus sp.				
ORGANISM	Mus sp.				
REFERENCE	1	(bases 1 to 2867)			
AUTHORS	Alonso,C.M., Domingo,D.G., Grandien,A., Leonardo,B. and Martinez,P.				
TITLE	Genes encoding for the human and murine death inducer-oblierator-1				

JOURNAL

Patent: JP 2002526040-A 2 20-AUG-2002;
 CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS

COMMENT

OS Mus sp. (murine)
 PN JP 2002526040-A/2
 PD 20-AUG-2002
 PF 10-SEP-1999 JP 2000570314
 PR 10-SEP-1998 SE 9803069-5,17-SEP-1998 US 60/100873 PI
 CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, PI
 ESTHER LEONARDO,
 PI PEDRO MARTINEZ
 PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC
 A61P29/00,
 PC A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC
 C12N5/10,
 PC C12P1/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC
 C12N15/00,
 PC C12N5/00,A61K37/02
 CC Genes encoding for the human and murine death inducer- CC
 obliterater-1
 FH Key Location/Qualifiers
 FT source 1..2867 /organism='Mus sp. (murine)'.
 Location/Qualifiers
 1..2867 /organism='Mus sp. (murine)'.
 /mol_type='genomic DNA'
 /db_xref='taxon:10095'

ORIGIN

Query Match 100.0%; Score 2867; DB 6; Length 2867;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2867; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES

source

Location/Qualifiers
 1..2867 /organism='Mus sp. (murine)'.
 /mol_type='genomic DNA'
 /db_xref='taxon:10095'

QY 1 TCCGTGAGTCTCTGGAATGCTGCGGATCCCGCGCGGAGCTTTGTTAAGG 60
 DB 1 TCCGTGAGTCTCTGGAATGCTGCGGATCCCGCGCGGAGCTTTGTTAAGG 60
 QY 61 CAGTCCCACTGTCCTTGTGTTTCCGAGCTCCGGAATCTTCAATTGAAGACTG 120
 DB 61 CAGTCCCACTGTCCTTGTGTTTCCGAGCTCCGGAATCTTCAATTGAAGACTG 120
 QY 121 ACTGCAACAGTGGGGTGAAGCTTGGCGCTGCTTGAAGCTTGGCCCAAGTCTATATT 180
 DB 121 ACTGCAACAGTGGGGTGAAGCTTGGCGCTGCTTGAAGCTTGGCCCAAGTCTATATT 180
 QY 121 ACTGCAACAGTGGGGTGAAGCTTGGCGCTGCTTGAAGCTTGGCCCAAGTCTATATT 180
 DB 121 ACTGCAACAGTGGGGTGAAGCTTGGCGCTGCTTGAAGCTTGGCCCAAGTCTATATT 180
 QY 181 TATGTAGATTTCCAGCCCAAGGTTTCAAGCTTTCAAGTTTGGGACAGTATGAT 240
 DB 181 TATGTAGATTTCCAGCCCAAGGTTTCAAGCTTTCAAGTTTGGGACAGTATGAT 240
 QY 241 AAAGGGACCTGAGCAATGAGAGCAACCAAGGCTATCAACCCAGTAAAGAGTTT 300
 DB 241 AAAGGGACCTGAGCAATGAGAGCAACCAAGGCTATCAACCCAGTAAAGAGTTT 300
 QY 301 AGGAAACCTGAGGTTTTCAGAGAACCAAGATGCAACCTGAGGGTGCAGAGACG 360
 DB 301 AGGAAACCTGAGGTTTTCAGAGAACCAAGATGCAACCTGAGGGTGCAGAGACG 360
 QY 361 GAGGGGACCCCAAGTGAAGCAACCAAGGCTTCAAGCTTCCCTGCGCGAGTGA 420
 DB 361 GAGGGGACCCCAAGTGAAGCAACCAAGGCTTCAAGCTTCCCTGCGCGAGTGA 420
 QY 421 CGGCAACCAAAACCTGAGAGGTTAGAGAGTTTCTTACAGAGGTTCCGCGCGAGG 480
 DB 421 CGGCAACCAAAACCTGAGAGGTTAGAGAGTTTCTTACAGAGGTTCCGCGCGAGG 480
 QY 481 AAAAAGAAATGTCCTGAGGATTCAGTGAAGCCCAATCTTCCAGATCACT 540
 DB 481 AAAAAGAAATGTCCTGAGGATTCAGTGAAGCCCAATCTTCCAGATCACT 540
 QY 541 GATGTGAGACAGCTTCCGAGGGGAGCTTGAAGCAATTTGAAATCAAGAGTGGCTT 600
 DB 541 GATGTGAGACAGCTTCCGAGGGGAGCTTGAAGCAATTTGAAATCAAGAGTGGCTT 600

QY 601 GTATCTGACTCTTAGAGAAAGACATCTGCTCTTCTGAAAGGCAAGAGCTGAA 660
 DB 601 GTATCTGACTCTTAGAGAAAGACATCTGCTCTTCTGAAAGGCAAGAGCTGAA 660
 QY 661 GAGGAAAGACACTCTGACAGTACAGTATGCTTAACTTGAAGAACTTCAAGC 720
 DB 661 GAGGAAAGACACTCTGACAGTACAGTATGCTTAACTTGAAGAACTTCAAGC 720
 QY 721 CGCCTTCCGAGAAACCGAGAGCAAGAACTGTGAGAGGCTCCCTGAGAGGCTGCAAT 780
 DB 721 CGCCTTCCGAGAAACCGAGAGCAAGAACTGTGAGAGGCTCCCTGAGAGGCTGCAAT 780
 QY 781 CGCCTGAGAGAAACCGAGAGCAAGAACTGTGAGAGGCTCCCTGAGAGGCTGCAAT 840
 DB 781 CGCCTGAGAGAAACCGAGAGCAAGAACTGTGAGAGGCTCCCTGAGAGGCTGCAAT 840
 QY 841 AGTCCGAGACAGACAGACTCTCTGTAAAGAGAGCTTGAAGCTTGAAGCAAGT 900
 DB 841 AGTCCGAGACAGACAGACTCTCTGTAAAGAGAGCTTGAAGCTTGAAGCAAGT 900
 QY 901 TCCCACTCAGACAGATGACATAGAAAATCAGTTGGAAGGAGGCGACTCAGAGGAAT 960
 DB 901 TCCCACTCAGACAGATGACATAGAAAATCAGTTGGAAGGAGGCGACTCAGAGGAAT 960
 QY 961 ACAGAGAAACCCCAAGGAGAGCGGCAACCAAGGCTGAGTGTGAGGTTTACAGCCC 1020
 DB 961 ACAGAGAAACCCCAAGGAGAGCGGCAACCAAGGCTGAGTGTGAGGTTTACAGCCC 1020
 QY 1021 AATGCTCTGTACTGCAATCTGCGCAAGCTTCAACCAAGGTTTATGATCTGTGAT 1080
 DB 1021 AATGCTCTGTACTGCAATCTGCGCAAGCTTCAACCAAGGTTTATGATCTGTGAT 1080
 QY 1081 CGCTGTGAGAGTGTGTTCAATGTGATCTGTGTGGTATTTTCAAGGCGGAGGCGGCTC 1140
 DB 1081 CGCTGTGAGAGTGTGTTCAATGTGATCTGTGTGGTATTTTCAAGGCGGAGGCGGCTC 1140
 QY 1141 CTGAAAGAAACGAGGAGAGACTACATCTGCGCAATTTGACCAATTTGCAAGTGCAGAT 1200
 DB 1141 CTGAAAGAAACGAGGAGAGACTACATCTGCGCAATTTGACCAATTTGCAAGTGCAGAT 1200
 QY 1201 GAGCAAAACGTTAGCGGCAACATGACAGAGCTTGGGTGCAATCTGTGGTGTGAT 1260
 DB 1201 GAGCAAAACGTTAGCGGCAACATGACAGAGCTTGGGTGCAATCTGTGGTGTGAT 1260
 QY 1261 GCGACAGCTGCAAGCAATGAGGACAGTGAAGCAAGTCCGAGAGAGCAAGGCGATA 1320
 DB 1261 GCGACAGCTGCAAGCAATGAGGACAGTGAAGCAAGTCCGAGAGAGCAAGGCGATA 1320
 QY 1321 AAGGTAGGATTTGAGAGGAGCAACCCAGCGGCAAGAAATCTCAAGATATTCCAG 1380
 DB 1321 AAGGTAGGATTTGAGAGGAGCAACCCAGCGGCAAGAAATCTCAAGATATTCCAG 1380
 QY 1381 CCTGTGTAGAGGCTCTGTGTCTCTTAAATGCAATTTGCGGCTTCTGAGTGAAG 1440
 DB 1381 CCTGTGTAGAGGCTCTGTGTCTCTTAAATGCAATTTGCGGCTTCTGAGTGAAG 1440
 QY 1441 CAGCTGACTGTGTATTTGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1441 CAGCTGACTGTGTATTTGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1500
 QY 1501 AGATTTCTAAGTTCAAGTAAAGCAAAACCAAGGAGGAGGAGGAGGAGGAGGAGG 1560
 DB 1501 AGATTTCTAAGTTCAAGTAAAGCAAAACCAAGGAGGAGGAGGAGGAGGAGGAGG 1560
 QY 1561 CCAAGAAAGTTCAAGTCTTCAAAATGCAAGTGTCAAGTGGGATTTAAATCTCTTCTG 1620
 DB 1561 CCAAGAAAGTTCAAGTCTTCAAAATGCAAGTGTCAAGTGGGATTTAAATCTCTTCTG 1620
 QY 1621 CACAAGAGACTAGGCTCAGAGAAAGGAGAAACCAAGTGAAGAAAGTGTGCTGCTCC 1680
 DB 1621 CACAAGAGACTAGGCTCAGAGAAAGGAGAAACCAAGTGAAGAAAGTGTGCTGCTCC 1680
 QY 1681 AGAGTGAAGCTTCTGGAGAGGAGGAGGCTGTGAGAGCAACATCTCTGCGCAAGT 1740

Db	1661	AGAGGTGAGACTTCGGAGAAAGGACACCTGTGAGACACACACACACTCTCGGACAGT	1740
Oy	1741	GACCAACAATCAATGCTGTGAAGCAGAGAAGCCAGAGAAGCCACTGCATCTCGCCC	1800
Db	1741	GACCAACAATCAATGCTGTGAAGCAGAGAAGCCAGAGAAGCCACTGCATCTCGCCC	1800
Oy	1801	ACCCATTATGAGTAAATGTACGTATCACCCAAAGCTGGCTTCCAGAGCTCCCATCAT	1860
Db	1801	ACCCATTATGAGTAAATGTACGTATCACCCAAAGCTGGCTTCCAGAGCTCCCATCAT	1860
Oy	1861	CTGGGTGGCTGCTCGGGGCTGTCTTAGGACACAGAGCTCGGGGTGTCTGGTGTCTATGTGA	1920
Db	1861	CTGGGTGGCTGCTCGGGGCTGTCTTAGGACACAGAGCTCGGGGTGTCTGGTGTCTATGTGA	1920
Oy	1921	GCACGACATCACTGCACGACAGAAAGCATATACCAAGATGCTCTGACCCCAAGTTC	1980
Db	1921	GCACGACATCACTGCACGACAGAAAGCATATACCAAGATGCTCTGACCCCAAGTTC	1980
Oy	1981	CTGCCTAGCCCTGTGGAGGCTCTCTGGGTGGTGTCTTAAAGACTGTGTAGGCTCATGTTG	2040
Db	1981	CTGCCTAGCCCTGTGGAGGCTCTCTGGGTGGTGTCTTAAAGACTGTGTAGGCTCATGTTG	2040
Oy	2041	GAGGCAATTTCTTATTTCAGTTTTAGGCTTGGTAAATTTGAATGCTCTCAACAGA	2100
Db	2041	GAGGCAATTTCTTATTTCAGTTTTAGGCTTGGTAAATTTGAATGCTCTCAACAGA	2100
Oy	2101	TGGGACACATCAGTGCATGTCAGATCAGATTCAGACTGAGTTCCTGCACAGACAGC	2160
Db	2101	TGGGACACATCAGTGCATGTCAGATCAGATTCAGACTGAGTTCCTGCACAGACAGC	2160
Oy	2161	CCAGCACAACCATGAGATTTTCCACGTTACTATTTTCTTCAACATGTCACACTG	2220
Db	2161	CCAGCACAACCATGAGATTTTCCACGTTACTATTTTCTTCAACATGTCACACTG	2220
Oy	2221	CTCTTCTATAGAAAAGTATTTTTTCAATGATTAGGCAATTTTGAAGAAGCAGGTAGA	2280
Db	2221	CTCTTCTATAGAAAAGTATTTTTTCAATGATTAGGCAATTTTGAAGAAGCAGGTAGA	2280
Oy	2281	CAATTGGTGTATTTGCATATCAACACCTACACATAACAGGTTTGAATTTGAAGGTC	2340
Db	2281	CAATTGGTGTATTTGCATATCAACACCTACACATAACAGGTTTGAATTTGAAGGTC	2340
Oy	2341	AATGCTGCTCAGATCCCATGGTGAATGCTGCTGCACATATACCTACAGCATATTC	2400
Db	2341	AATGCTGCTCAGATCCCATGGTGAATGCTGCTGCACATATACCTACAGCATATTC	2400
Oy	2401	TCATCAAGTGAGAAATGTGGGTGTAGCTTCCAGACTCGTAGATGCCAGATTTCAA	2460
Db	2401	TCATCAAGTGAGAAATGTGGGTGTAGCTTCCAGACTCGTAGATGCCAGATTTCAA	2460
Oy	2461	TCAGTGTATTATGTTTGAATTAATCTTTGAAACCCAGATTAAGTCACTAGATTGAGC	2520
Db	2461	TCAGTGTATTATGTTTGAATTAATCTTTGAAACCCAGATTAAGTCACTAGATTGAGC	2520
Oy	2521	CATTCTGTTCATCTGCTGACAAAGTACACTTCTTCACTTTTGTGAACAGTGGTTCAA	2580
Db	2521	CATTCTGTTCATCTGCTGACAAAGTACACTTCTTCACTTTTGTGAACAGTGGTTCAA	2580
Oy	2581	GTTGAATTTTCACTGTGTGTTCTCATGATATGTTTGTTCATCTCTGTCCACCTCAGCC	2640
Db	2581	GTTGAATTTTCACTGTGTGTTCTCATGATATGTTTGTTCATCTCTGTCCACCTCAGCC	2640
Oy	2641	CCAAGTAAAGGGGTCTCATGGCTTAAGCTGGCCCTTAAATTTTCTGTAAAGTTGGGGTTC	2700
Db	2641	CCAAGTAAAGGGGTCTCATGGCTTAAGCTGGCCCTTAAATTTTCTGTAAAGTTGGGGTTC	2700
Oy	2701	TCTGATCCCACTGTAAAGAGTGTCTGAATCTTACAGGACACACACTGTATATGCTGTG	2760
Db	2701	TCTGATCCCACTGTAAAGAGTGTCTGAATCTTACAGGACACACACTGTATATGCTGTG	2760
Oy	2761	GGGTGTGAATCAAGATGTGTGAACCAAGACTCTACAGCTGAGCTGACGTTCTAGCCATG	2820

Db	2761	GGGGTGAAGTCAAGAGTGTGGACACAGAGCTTACAGACGTGAGCTGCAGTTCTAGCATG	28252
QY	2821	GTAACGTAACTGCTGATCTTCTTGTCTTTGTTCCCTTGTGCTGG	2867
Db	2821	GTAACTGAACTGCTGATCTTCTTGTCTTTGTTCCCTTGTGCTGG	2867
RESULT 2			
AX023370			
LOCUS	AX023370	2867 bp	DNA
DEFINITION	Sequence 2 from Patent W00015787.		linear
ACCESSION	AX023370		
VERSION	AX023370.1	GI:10183782	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match	100.0%;	Score 2867;	DB 6; Length 2867;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 2867;	Conservative	0; Mismatches	0; Indels 0; Gaps 0;
QY	1	TCCTGTGTAGCTGTGAAATGTGCTGCGGATCCCGGCGCCGGGAGCTTGTAAAGG	60
Db	1	TCCTGTGTAGCTGTGAAATGTGCTGCGGATCCCGGCGCCGGGAGCTTGTAAAGG	60
QY	61	CAGTCCCCACGTGTCCCTTGTGTGTGTGGAAGCTCCGGAATCTTCTCATGTATGAACTGTG	120
Db	61	CAGTCCCCACGTGTCCCTTGTGTGTGTGGAAGCTCCGGAATCTTCTCATGTATGAACTGTG	120
QY	121	ACTGACAAACGTGGGGGTGAGGCTTGCGCGATGTGCTTGAACCTGCGCCGACAGGCTTAATTT	180
Db	121	ACTGACAAACGTGGGGGTGAGGCTTGCGCGATGTGCTTGAACCTGCGCCGACAGGCTTAATTT	180
QY	181	TATGTAGGATTTCCAGCCAAAGGTTTCAAGCTTTCAGTTTGTGGACAGGTATGATGAT	240
Db	181	TATGTAGGATTTCCAGCCAAAGGTTTCAAGCTTTCAGTTTGTGGACAGGTATGATGAT	240
QY	241	AAAGGGCACTGAGCAATGAGGAAGCAACCAAGGCTATCAAACTCAGTAAAGAGTTT	300
Db	241	AAAGGGCACTGAGCAATGAGGAAGCAACCAAGGCTATCAAACTCAGTAAAGAGTTT	300
QY	301	AGGAAACCTGTGGGTTTTGAAAGAACCAAGATTTGCCAAAGTGAAGGTGACAGACACG	360
Db	301	AGGAAACCTGTGGGTTTTGAAAGAACCAAGATTTGCCAAAGTGAAGGTGACAGACACG	360
QY	361	GAGCGGACCCCGAGTGAAGCAACCAAGATTTGCCAAAGTGAAGGTGACAGACACG	420
Db	361	GAGCGGACCCCGAGTGAAGCAACCAAGATTTGCCAAAGTGAAGGTGACAGACACG	420
QY	421	CGGCAACCAAAAGTACCTGAGAGAGGTGAAGAGTTTCTTACCAAGGTTCCGCGCCGAGG	480
Db	421	CGGCAACCAAAAGTACCTGAGAGAGGTGAAGAGTTTCTTACCAAGGTTCCGCGCCGAGG	480
QY	481	AAAAAGAAATGTGCGCGGTGTCCCTGAGAGGATTTCCAGTGAAGCCCATCTTCCACAGTACT	540
Db	481	AAAAAGAAATGTGCGCGGTGTCCCTGAGAGGATTTCCAGTGAAGCCCATCTTCCACAGTACT	540

QY 541 GATGTGAGACAGCTTCCGAGGGAGCGTGAAGACAGTTCTGAGATCAGAGTGGCCCT 600
DB 541 GATGTGAGACAGCTTCCGAGGGAGCGTGAAGACAGTTCTGAGATCAGAGTGGCCCT 600
QY 601 GTATCTGACTCTTATGAGGAAAGAAATCTCTGCTCTTCTGAAAAGGCAAAAGAGGTAA 660
DB 601 GTATCTGACTCTTATGAGGAAAGAAATCTCTGCTCTTCTGAAAAGGCAAAAGAGGTAA 660
QY 661 GAGGAAGAGACACTCTGACAGTGAAGTGGCTTACGTTGAGAGAACTTCAAGAC 720
DB 661 GAGGAAGAGACACTCTGACAGTGAAGTGGCTTACGTTGAGAGAACTTCAAGAC 720
QY 721 CGCTTTCGGAAGAACCGGACGACAGAACTTCTGAGAGAGTCCCTGAGAGGACAGTGAAT 780
DB 721 CGCTTTCGGAAGAACCGGACGACAGAACTTCTGAGAGAGTCCCTGAGAGGACAGTGAAT 780
QY 781 CGCTTTCGGAAGAACCGGACGACAGAACTTCTGAGAGAGTCCCTGAGAGGACAGTGAAT 840
DB 781 CGCTTTCGGAAGAACCGGACGACAGAACTTCTGAGAGAGTCCCTGAGAGGACAGTGAAT 840
QY 841 AGTCCGAGACAGACAGACTCTCTGTAAGCAGAGCTTGAAGGCTTGAAGGACAGTGAAT 900
DB 841 AGTCCGAGACAGACAGACTCTCTGTAAGCAGAGCTTGAAGGCTTGAAGGACAGTGAAT 900
QY 901 TCCCACTGAGAGACAGATGACATGAAATCACTTGAAGGAGGACGACACTCAAGGAAAT 960
DB 901 TCCCACTGAGAGACAGATGACATGAAATCACTTGAAGGAGGACGACACTCAAGGAAAT 960
QY 961 ACAGAGGAAACCCCGAGGAGACGAGCAAAACCAAGGCTGAGTGAAGGTTTACAGCC 1020
DB 961 ACAGAGGAAACCCCGAGGAGACGAGCAAAACCAAGGCTGAGTGAAGGTTTACAGCC 1020
QY 1021 AATGCTCTGTAAGTGAATCTGCTGCAAGCTTCAACAAACAGGTTTATGATCTGCTGAT 1080
DB 1021 AATGCTCTGTAAGTGAATCTGCTGCAAGCTTCAACAAACAGGTTTATGATCTGCTGAT 1080
QY 1081 CGGTGTGAGAGTGGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 CGGTGTGAGAGTGGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 CTGGAAGGAAACCGGAGAGACATCACTGCTGCAAAATTTGCAACATTTGCAAGTGAAT 1200
DB 1141 CTGGAAGGAAACCGGAGAGACATCACTGCTGCAAAATTTGCAACATTTGCAAGTGAAT 1200
QY 1201 GAGACAAACCGTGAAGGCAACATGAGAGAGACTCTGAGTGAATCTGAGTGAAT 1260
DB 1201 GAGACAAACCGTGAAGGCAACATGAGAGAGACTCTGAGTGAATCTGAGTGAAT 1260
QY 1261 GGCACAGCTGACAGAGCATGAGGACAGTGAAGAGAGTCCGAGAGAGACAGAGGACAT 1320
DB 1261 GGCACAGCTGACAGAGCATGAGGACAGTGAAGAGAGTCCGAGAGAGACAGAGGACAT 1320
QY 1321 AAGGTGAAGTGAAGAGCAAAACCCAGCGGACAGAGAAATCTCAAGATATTCCAG 1380
DB 1321 AAGGTGAAGTGAAGAGCAAAACCCAGCGGACAGAGAAATCTCAAGATATTCCAG 1380
QY 1381 CCGTGTGAGAGGCTCTGAGTCTCTTAATGATGATGATGATGATGATGATGATGATGAT 1440
DB 1381 CCGTGTGAGAGGCTCTGAGTCTCTTAATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 CAGCTGAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1441 CAGCTGAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1501 AGATTTCTAAGTTCAGGTAAAGCAAAACCAAGGAAAGGTCAAGAGAG 1560
DB 1501 AGATTTCTAAGTTCAGGTAAAGCAAAACCAAGGAAAGGTCAAGAGAG 1560
QY 1561 CAGAGAAAGTTCAGGTTCAGAAATGCAAGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 1620
DB 1561 CAGAGAAAGTTCAGGTTCAGAAATGCAAGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 1620
QY 1621 CAGAGAAAGTTCAGGTTCAGAAATGCAAGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 1680
DB 1621 CAGAGAAAGTTCAGGTTCAGAAATGCAAGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 1680

DB 1621 CAGAGAAAGTTCAGGTTCAGAAATGCAAGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 1680
QY 1681 AGAGTGAAGCTTCTGAGAGAGAGGACACTCTGTAAGACAGACATCTCGGGACAGT 1740
DB 1681 AGAGTGAAGCTTCTGAGAGAGAGGACACTCTGTAAGACAGACATCTCGGGACAGT 1740
QY 1741 GAGCAACACTCAATGCTGTAAGGACAGAGAGAGGACAGAGGACAGTCTGAGCC 1800
DB 1741 GAGCAACACTCAATGCTGTAAGGACAGAGAGAGGACAGAGGACAGTCTGAGCC 1800
QY 1801 AACCATTGAGTAAATGTAAGTATCAACCAAGGCTGCTTCCAGGCTTCCATCAT 1860
DB 1801 AACCATTGAGTAAATGTAAGTATCAACCAAGGCTGCTTCCAGGCTTCCATCAT 1860
QY 1861 CTGGGTGCTGCTGAGGCTGCTGTAAGACAGAGGCTGAGGCTGAGTCTGAGTGAAT 1920
DB 1861 CTGGGTGCTGCTGAGGCTGCTGTAAGACAGAGGCTGAGGCTGAGTCTGAGTGAAT 1920
QY 1921 GGCACAGCTCACTGACAGGACAGAGACAGATACCAAGTGCCTCTGAGACCCAGGTTC 1980
DB 1921 GGCACAGCTCACTGACAGGACAGAGACAGATACCAAGTGCCTCTGAGACCCAGGTTC 1980
QY 1981 CTGCTGAGCTGAGAGCTCTCTGAGGCTTCTTAAGAGCTGTAAGGCTCAATGTTG 2040
DB 1981 CTGCTGAGCTGAGAGCTCTCTGAGGCTTCTTAAGAGCTGTAAGGCTCAATGTTG 2040
QY 2041 GAGGCAATTTCTAATTTAGGCTTGAAGGCTTGAAGTATTTGAAGTCTCAACAGA 2100
DB 2041 GAGGCAATTTCTAATTTAGGCTTGAAGGCTTGAAGTATTTGAAGTCTCAACAGA 2100
QY 2101 TGGACAAACATGAGTCACTGCTGAGTCAAGTCAAGTTCCTCAAGAGCAGC 2160
DB 2101 TGGACAAACATGAGTCACTGCTGAGTCAAGTCAAGTTCCTCAAGAGCAGC 2160
QY 2161 CAGACAAACATGAGTTCACAGTTCCTGAGTCAATTTCTCAACAGTCAACAGC 2220
DB 2161 CAGACAAACATGAGTTCACAGTTCCTGAGTCAATTTCTCAACAGTCAACAGC 2220
QY 2221 CTTCTCTAAGAAAGTATTTTTCATGAGTGAAGGCAATTTGAAGAGCAGGTAGA 2280
DB 2221 CTTCTCTAAGAAAGTATTTTTCATGAGTGAAGGCAATTTGAAGAGCAGGTAGA 2280
QY 2281 CAATTTGCTAATTTCACTACACACTACATTAACAGGTTGAATTTGAAGGTC 2340
DB 2281 CAATTTGCTAATTTCACTACACACTACATTAACAGGTTGAATTTGAAGGTC 2340
QY 2341 ATCTGCTCTCAGATCCCATGAGTGAATCTGCTGACATACACTACAGAGATATTC 2400
DB 2341 ATCTGCTCTCAGATCCCATGAGTGAATCTGCTGACATACACTACAGAGATATTC 2400
QY 2401 TCAATCAAGTGAAGTATGAGTGAAGTTCACATCTGTAAGTCCCAAGATTTCAAA 2460
DB 2401 TCAATCAAGTGAAGTATGAGTGAAGTTCACATCTGTAAGTCCCAAGATTTCAAA 2460
QY 2461 TCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2520
DB 2461 TCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2520
QY 2521 CATCTGTTTCACTGCTGTAAGGTAACCTTTCTTAATTTTGTACAGTGGTTCAA 2580
DB 2521 CATCTGTTTCACTGCTGTAAGGTAACCTTTCTTAATTTTGTACAGTGGTTCAA 2580
QY 2581 GGAATTTTCACTGCTGTAAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2640
DB 2581 GGAATTTTCACTGCTGTAAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2640
QY 2641 CAGAGTAAAGGAGTCTCAATGAGCTGAGCTTAAATTTCTGTAAGTGGGAGTCC 2700
DB 2641 CAGAGTAAAGGAGTCTCAATGAGCTGAGCTTAAATTTCTGTAAGTGGGAGTCC 2700
QY 2701 TCTGCTCCCACTGTAAGAGTGTGAATCTTAAGGAGCAGACCTGATGATGCTGAG 2760
DB 2701 TCTGCTCCCACTGTAAGAGTGTGAATCTTAAGGAGCAGACCTGATGATGCTGAG 2760

Oy	2761	GCGGTGTAACCTCAGAGTGTGGACACAGTCTAACAGCTGAGCTGCAGTTTAGCCATG	2820
Db	2761	GCGGTGTAACCTCAGAGTGTGGACACAGTCTAACAGCTGAGCTGCAGTTTAGCCATG	2820
Oy	2821	GTAACTGAACACTGCTGATCTTCTTGCTCTTTGTTTCCTTTGTCGCGG	2867
Db	2821	GTAACTGAACACTGCTGATCTTCTTGCTCTTTGTTTCCTTTGTCGCGG	2867
RESULT 3			
LOCUS	MWU238332	2602 bp	mRNA linear ROD 09-JUN-1999
DEFINITION	Mus musculus mRNA for death inducer-obliterator-1 (Dio-1).		
ACCESSION	AJ238332		
VERSION	AJ238332.1 GI:54577402		
KEYWORDS	death inducer-obliterator-1; Dio-1 gene.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	1 Garcia-Domingo,D., Leonardo,E., Grandien,A., Martinez,P., Albar,J.P., Izpisua-Belmonte,J.C. and Martinez-A., C. DIO-1 is a gene involved in onset of apoptosis in vitro, whose misexpression disrupts limb development Proc. Natl. Acad. Sci. U.S.A. 96 (14), 7992-7997 (1999) 99324176 10393935		
JOURNAL	2 (bases 1 to 2602)		
PUBLISHED	Garcia-Domingo, D.		
REFERENCE	Direct Submission		
TITLE	Submitted (21-APR-1999) Garcia-Domingo D., Department of Immunology and Oncology, Centro Nacional de Biotecnologia/CNIC, Campus Candablanco UAM, Ctra. de Colmenar Viejo Km 16, Madrid, 28049, SPAIN		
JOURNAL			
FEATURES			
source	location/Qualifiers		
	1..2602		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/cell_line="MOL-1"		
	/cell_type="pre-BI"		
	1..2602		
	/gene="Dio-1"		
	232..>2073		
	/gene="Dio-1"		
	/function="apoptosis inducer"		
	/codon_start=1		
	/evidence=experimental		
	/product="death inducer-obliterator-1"		
	/protein_id="CA848401.1"		
	/db_xref="GI:5457403"		
	/db_xref="GOA:Q8C9B9"		
	/db_xref="GOA:Q9WV00"		
	/db_xref="SMIS-PROT:Q8C9B9"		
	/translation="MDKGLHSNEBPALKIPTSKEPKTKMGPRRTTIARBGSGTSE ADPSBOOPQOHNLSIRSGOGPDKTRVEVFLLTVRRGKNPPVSLSDSSBPSTSV TDVETLRSRGVSSESSRLTSGDPVDSLSKHPASERKXGGBEEEDTSDSDGLTLKE LNLRNRKRBOEPYERSLRGSQPNLRKKREBEESALRGSVQIGAEQDRPLCKQPEPA SQGPQSOSFETDIENOCLECATGNTENPREAGKPCECVDPNALYCICQPENNA RFPIICDCRGEHFHGDVCVIGSEARGLLBNBGDIYCPCNLTQLQVDETNGSATNBOD SGCRSYAGDGTCTGCTGTVRKSGSDGIGRIERKANPGCKKKIKFOGVVARCAP KCIKGPCSSVAOPDPYSKNDCLIKHAAIMNLPLSSGEKOKTPKERVTKPERKPSLP KCSVOVGKISSVHKRLASKRENPKVMKLARSSETSGKBACESTPTWSADHNTN AVNPKRPKPTALSPFLSKCTYHPKAGFGPGSHLDGCLGRTRLVLGLVLIIVASS SLPARSRYPQASGPQVFLPILMSLSGVTLSGVLMLEAISYFSFRPW"		
ORIGIN			
Query Match	90.8%; Score 2602;	DB 10;	Length 2602;
Best Local Similarity	100.0%;	Fred. No. 0;	
Matches 2602;	Conservativity 0;	Mismatches 0;	Indels 0;
		Gaps 0;	

QY	1	TCGCTGTAGCTCTGGAAATGAGCTGCGGGATCCCGCGCGCCCGGGAGCTTGTTTAAGG	60
Db	1	TCGCTGTACTCTGGAAATGAGCTGCGGGATCCCGCGCGCCGGGAGCTGTTAAGG	60
QY	61	CAGTCCCCCACTGCTCCCTTGGTGTGTGAAAGCTCCGGAAATCTTCTCATGTATGAATCTG	120
Db	61	CAGTCCCCCACTGTCTCTTGGTGTGTGAAAGCTCCGGAAATCTTCTCATGTATGAATCTGTG	120
QY	121	ACTGACAAACATGAGGGGTGAGGCTTGCGCGTCTGCTTGACTGCGCCCAAGTCTATAATT	180
Db	121	ACTGACAAACATGAGGGGTGAGGCTTGCGCGTCTGCTTGACTGCGCCCAAGTCTATAATT	180
QY	181	TATGTAGAAATTTCCAGGCAAAAGGTTTCCAAAGCTTCAAGTTTGAGTTTGGACAGGTATGAT	240
Db	181	TATGTAGAAATTTCCAGGCAAAAGGTTTCCAAAGCTTCAAGTTTGGACAGGTATGAT	240
QY	241	AAAGGCACTGTAGCAATGAGAAAGCACCCAAGGCTATCAAAACCACAGTAAAGAGTTT	300
Db	241	AAAGGCACTGTAGCAATGAGAAAGCACCCAAGGCTATCAAAACCACAGTAAAGAGTTT	300
QY	301	AGGAAACCTGAGGTTTTCGAAAGAACCAATTCGCAACCTGAGAGGTGAGAGACACG	360
Db	301	AGGAAACCTGAGGTTTTCGAAAGAACCAATTCGCAACCTGAGAGGTGAGAGACACG	360
QY	361	GAGCGGACCCCAAGTGAAGACACACAGAGCATAACTCTCCCTGCGCGCAGTGA	420
Db	361	GAGCGGACCCCAAGTGAAGACACACAGAGCATAACTCTCTCCCTGCGCGCAGTGA	420
QY	421	CGGACCAAAACGTACTGAGAGGGTGAAGAGTTTCTTACACAGGTTCCGCGCCGAGGG	480
Db	421	CGGACCAAAACGTACTGAGAGGGTGAAGAGTTTCTTACACAGGTTCCGCGCCGAGGG	480
QY	481	AAAAAGATGTGCCGAGTCCCTGGAAGATTCAGATGAGGCCACATCTTCCACAGTCACT	540
Db	481	AAAAAGATGTGCCGAGTCCCTGGAAGATTCAGATGAGGCCACATCTTCCACAGTCACT	540
QY	541	GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGACAGTTCAGATCAGAAGTGACCT	600
Db	541	GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGACAGTTCAGATCAGAAGTGACCT	600
QY	601	GTATCTGACTCTTAGGAAAGAAACATCTGCTCTTCTGAAAGGCAAAAGAGGTGA	660
Db	601	GTATCTGACTCTTAGGAAAGAAACATCTGCTCTTCTGAAAGGCAAAAGAGGTGA	660
QY	661	GAGGAAGAAGACCTCTGACAGTGAAGTATGGCTTACGTTGAAAGAACTTCAGAAC	720
Db	661	GAGGAAGAAGACCTCTGACAGTGAAGTATGGCTTACGTTGAAAGAACTTCAGAAC	720
QY	721	CGCTTCTGAGAAAGCGAGACAGAACTGTGAGAGGTCTCTGAGAGCGATCAGAT	780
Db	721	CGCTTCTGAGAAAGCGAGACAGAACTGTGAGAGGTCTCTGAGAGCGATCAGAT	780
QY	781	CGCTGTGAGAAAGCGCAGAGGAAAGATTTGCGCAAACTGGGAGGTCAATATGGC	840
Db	781	CGCTGTGAGAAAGCGCAGAGGAAAGATTTGCGCAAACTGGGAGGTCAATATGGC	840
QY	841	AGTGCACGACGACAGACACTCTCTGTAAACAGAGCCTGAGGCTAGTCAAGGACCACTG	900
Db	841	AGTGCACGACGACAGACACTCTCTGTAAACAGAGCCTGAGGCTAGTCAAGGACCACTG	900
QY	901	TCCCACTGACAGACAGATGACATRGAATACTAGTTGAAAGGAAAGGCACTCAAGGAAAT	960
Db	901	TCCCACTGACAGACAGATGACATRGAATACTAGTTGAAAGGAAAGGCACTCAAGGAAAT	960
QY	961	ACAGAGGAAACCCCAAGGAAAGCGGCAAAACAAAGCCTGAGTGTAGAGTTTACACCCC	1020
Db	961	ACAGAGGAAACCCCAAGGAAAGCGGCAAAACAAAGCCTGAGTGTAGAGTTTACACCCC	1020
QY	1021	AATGCTCTGTATCTGCATCTGCGCGCAGCTTCAACAACAGGTTTATGATCTGCTGTAT	1080
Db	1021	AATGCTCTGTATCTGTCTGCGCGCAGCTTCAACAACAGGTTTATGATCTGCTGTAT	1080

QY 1081 CGGTGTGAGAGAGTGTTCATATGTCATCTGTGTGGTATTTCTGAGGCCGAGGCGGCTC 1140
 DB 1081 CGGTGTGAGAGAGTGTTCATATGTCATCTGTGTGGTATTTCTGAGGCCGAGGCGGCTC 1140
 QY 1141 CTGGAAGAGAGAGGAGAGAGCTATGCTGCGCAAAATTCACCATTTTTCAGAGTGCAGAT 1200
 DB 1141 CTGGAAGAGAGAGGAGAGAGCTATGCTGCGCAAAATTCACCATTTTTCAGAGTGCAGAT 1200
 QY 1201 GAGCAAAACGGTAGCGGCCCAATGAGCAGAGACTCTGTGGTGCAGATCTGTGGTGTCTGAT 1260
 DB 1201 GAGCAAAACGGTAGCGGCCCAATGAGCAGAGACTCTGTGGTGTCTGATCTGTGGTGTCTGAT 1260
 QY 1261 GGCACAGATCTGACAAAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 DB 1261 GGCACAGATCTGACAAAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1321 AAGGAGTATGAG 1380
 DB 1321 AAGGAGTATGAG 1380
 QY 1381 CCTGTGTGAGAGGCTCTGTGTCTCTAATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1381 CCTGTGTGAGAGGCTCTGTGTCTCTAATGATGATGATGATGATGATGATGATGATGATGAT 1440
 QY 1441 CAGCTTGAATCTGTGTATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1441 CAGCTTGAATCTGTGTATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 QY 1501 AGATTTCTAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 1560
 DB 1501 AGATTTCTAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 1560
 QY 1561 CCAGAAAAGTTCACTCTTCGAAATGCACTGTTTCAGTGTGGGATTTAAATCTCTCTGTG 1620
 DB 1561 CCAGAAAAGTTCACTCTTCGAAATGCACTGTTTCAGTGTGGGATTTAAATCTCTCTGTG 1620
 QY 1621 CACAAAGACTAGGCTCAG 1680
 DB 1621 CACAAAGACTAGGCTCAG 1680
 QY 1681 AGAGTGTGAGACTCTCTGAG 1740
 DB 1681 AGAGTGTGAGACTCTCTGAG 1740
 QY 1741 GACCAACAATCAATGCTGTGAG 1800
 DB 1741 GACCAACAATCAATGCTGTGAG 1800
 QY 1801 ACCCTATGAGTAAATGTAACCTATGTAACCTATGTAACCTATGTAACCTATGTAACCTATGTA 1860
 DB 1801 ACCCTATGAGTAAATGTAACCTATGTAACCTATGTAACCTATGTAACCTATGTAACCTATGTA 1860
 QY 1861 CTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 DB 1861 CTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 QY 1921 GGCACAGAGCTCACTGCGCAG 1980
 DB 1921 GGCACAGAGCTCACTGCGCAG 1980
 QY 1981 CTGCTTACGCTGTGAGAGCTCTCTGTGTGTGTCTTAAAGAGCTGTGTGTGTGTGTGTGTGTGT 2040
 DB 1981 CTGCTTACGCTGTGAGAGCTCTCTGTGTGTGTCTTAAAGAGCTGTGTGTGTGTGTGTGTGTGT 2040
 QY 2041 GAGGCAATTTCTTATTTAGCTTTAGGCTTTGTAAGTATTTGAACTGTGCTTCAACAGA 2100
 DB 2041 GAGGCAATTTCTTATTTAGCTTTAGGCTTTGTAAGTATTTGAACTGTGCTTCAACAGA 2100
 QY 2101 TGGGACAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 DB 2101 TGGGACAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 QY 2161 CGAGCAAAACATGAGATTTTCCAGGTTACTATTTTCTTCAACATGATGATGATGATGATGATGAT 2220

DB 2161 CCAGCAGAAACATGAGATTTTCCAGGTTACTATTTTCTTCAACATGATGATGATGATGATGATGAT 2220
 QY 2221 CTCTTCTATAGAAAAGATTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 DB 2221 CTCTTCTATAGAAAAGATTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 QY 2281 CAATTTGTCTATTTTCACTATGACACCTATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
 DB 2281 CAATTTGTCTATTTTCACTATGACACCTATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
 QY 2341 ATCTGTGCTCAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 DB 2341 ATCTGTGCTCAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 QY 2401 TCATCAAAAGTAAAGATTTGTGGT 2460
 DB 2401 TCATCAAAAGTAAAGATTTGTGGT 2460
 QY 2461 TCAGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 DB 2461 TCAGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 QY 2521 CATCTTGTTCATGCTGTGACAGAGTACACCTTTCTTACTTTTGTGACAGTGTTCAA 2580
 DB 2521 CATCTTGTTCATGCTGTGACAGAGTACACCTTTCTTACTTTTGTGACAGTGTTCAA 2580
 QY 2581 GTGAATTTTCACTGTGTGTCTC 2602
 DB 2581 GTGAATTTTCACTGTGTGTCTC 2602

RESULT 4

AKI29117
 LOCUS AKI29117 4906 bp mRNA linear ROD 21-NOV-2003
 DEFINITION Mus musculus mRNA for mKIAA0333 protein.

ACCESSION AKI29117
 VERSION AKI29117.1 GI:37359897
 KEYWORDS FLI_CDNA.
 SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

TITLE

1 Okazaki, N., Kihuno, R., Ohara, R., Inamoto, S., Koseki, H., Hiraoaka, S.,
 Sage, Y., Nagase, T., Ohara, O. and Koga, H.
 Prediction of the coding sequences of mouse homologues of KIAA
 gene: III. the complete nucleotide sequences of 500 mouse
 KIAA-homologous cDNAs identified by screening of terminal sequences
 of cDNA clones randomly sampled from size-fractionated libraries

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (23-JUL-2003) Hisashi Koga, Kazusa DNA Research
 Institute, Laboratory for Genome Informatics; 2-6-7
 Kazusa-kametari, Kisarazu, Chiba 292-0818, Japan
 (E-mail: mousekazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
 The GREAT program supported by Japan science and technology
 corporation; cDNA full insert sequencing; Kazusa DNA Research
 Institute; cDNA library construction, clone selection and 5'- &
 3'-end one pass sequencing.

FEATURES

source

1..4906
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="mbg16015"
 /tissue_type="brain"
 /dev_stage="adult"
 /note="vector:modified pBC SK+"

gene
CDS

```

1. .4906
/gene="mKIA0333"
<224. .3829
/gene="mKIA0333"
/Note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not experimental
/product="mKIA0333 protein"
/protein_id="BAC97927.1"
/db_xref="GI:37359898"
/translation="FVGPAPGAPAFSTGTGMDKGHLSNEAPALIPSTKEPKFT
MGRRTTIAKRGAGDTEVDPSSQOPOQNLILRSGRPKTRERVEPLTVRRGK
KAPVPLSDSSPTSDTVDVETASGSSVSSISGVSILSGHNSLSERKAGS
BEERBTDSDGLTLKELONLRREKROEVPVRLRKRREDSVETSV
OIGSABODPLCKOEPBAGVPSOSTDINOLEGKXNONTENPBAKREC
EYDPRALXICRQPHNRFNCCDCCEPHDCTGISIBARLLERGEDYICNC
TLYQVDETNSATBDSDSCRSVGADGTCSTIGVEQSGDQIKGIRKANPS
GRRKIKIPQVVPBAPKPCIGPCSSVAPDSVCSNDCILKHAATVPLSGKEQ
KTRPKKVTKEBESLPCKSVOVGKISVHKLBSRGRNPKVLMASLETGK
EACSSSTPSMASDHNVAVKPEKPKPTALSPITLSKMKDRAVEDTMAVTLPK
KALPSASLVGROTSPRNLYPKLIPYSNMAKAPAIKKLPSGKGTIPRPMPSATLS
GTSAAQAGTPHTTAAKSLPGSAAVYGTTRKPMASAVPAAPKGLIPVSPAPOSN
SIRKONIRSLKSLIMRVNDSDLLMTNBEVGLALIEKEMFNLPOYTDNRKSKY
KSLMFNLKDPKNQGLFHRVLARBEISLAKVNRMPKPELVKESLMTKEPTKYSISRT
RLINSSKNKTTPETIPDMEDEPVSDBEQDSVAAPEKSNAPLLDVFSSKLDIT
SGHRAHLPFNCKICTGQVSSSEDBAPAKQKLSASKKEDFPRSDSPNVAVPTA
DEGIADTPEKNSSEDPBSTSLNORKCFPSRSPDSHPESLGLSPSSAGSGGV
VTTVMGSDPPTALSGCTVYASMAHLKDNQOASPTKLDMLKALTSVAPKSLAK
PSSSDPRLTSPSPSISESRSPPEGDITLPLSRANTIKGPIINQSLAKPTTAKY
VSGCDYLSIEDLPDTHIGRLAPKTVMDYVGLKSVKELCLLIRHATBEERAY
ISLVSYSFSGRGRGVANNRHYKDYLLPLSAKDPVPSKLPFBGPGKPVSGR"

```

ORIGIN

Query Match 61.8%; Score 1770.4; DB 10; Length 4906;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1774; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

44 GGAGCTGTTTAAAGAGCAGTCCCACTGTCCTTGTGTGTGCAAGCTCCGAATCTT 103
90 GAGCTGTTTAAAGAGCAGTCCCACTGTCCTTGTGTGTGCAAGCTCCGAATCTT 149
104 CTCAATTGATGAACCTGTACTGCAACAGTGGGGGTGAGGCTTGGCCCTTGTGACCTGG 163
150 CTCAATTGATGAACCTGTACTGCAACAGTGGGGGTGAGGCTTGGCCCTTGTGACCTGG 209
164 CCCGAGGTCTAATATTTTATGTAGATTTCCAGCCCAAGGTTTCCAGCTTTCAGATTTTG 223
210 CCCGAGGTCTAATATTTTATGTAGATTTCCAGCCCAAGGTTTCCAGCTTTCAGATTTTG 269
224 GGAAGGTATGATGATTAAGGCACTTGAAGCAATGAGAGCAACCCCAAGGCTATCAAC 283
270 GGAAGGTATGATGATTAAGGCACTTGAAGCAATGAGAGCAACCCCAAGGCTATCAAC 329
284 CCACCAAGTAAGAGTTCAAGAAAACTTGGGGTTTTCGAAGAACCAAGATTTGCCAAACGTG 343
330 CCACCAAGTAAGAGTTCAAGAAAACTTGGGGTTTTCGAAGAACCAAGATTTGCCAAACGTG 389
344 AGGGTCAGAGAACACAGGAGGCGGACCCCAAGTGAAGAGAACCAAGAGCAACCACTCT 403
390 AGGGTCAGAGAACACAGGAGGCGGACCCCAAGTGAAGAGAACCAAGAGCAACCACTCT 449
404 CCCTGCGCGCAGTGAACGCAACCAAAAGTACTGAAGAGGTGAAGAGTTCTTAACA 463
450 CCCTGCGCGCAGTGAACGCAACCAAAAGTACTGAAGAGGTGAAGAGTTCTTAACA 509
464 CGGTTTCGGCGCGCAGAGGAAAAAAGATGTGCGCGGTGTCCTTGAAGATTCAGTGAAGCCCA 523
510 CGGTTTCGGCGCGCAGAGGAAAAAAGATGTGCGCGGTGTCCTTGAAGATTCAGTGAAGCCCA 569
524 CATCTTTCACATCTGATGTGTGAGACAGGCTTCCAGAGGAGAGCGTTGAAGAGAGTTCTTG 583
570 CATCTTTCACATCTGATGTGTGAGACAGGCTTCCAGAGGAGAGCGTTGAAGAGAGTTCTTG 629

```

```

584 AGATCAGAAAGTGGCCCTGTATCTGACTCTTTAGGAAAAAGAACTCTGCTCTTCTGA 643
630 AGATCAGAAAGTGGCCCTGTATCTGACTCTTTAGGAAAAAGAACTCTGCTCTTCTGA 689
644 AGGCAAAAAGAGGTGAAGAGAAAGAAAGCACTCTGAAGTGAAGAGAGGCTTAAGT 703
690 AGGCAAAAAGAGGTGAAGAGAAAGAAAGCACTCTGAAGTGAAGAGAGGCTTAAGT 749
704 TGAAGAACTTCAAGAACCGCTTTCGAGAAAGCGAGAGCAAGAACTGTGAGAGTCCC 763
750 TGAAGAACTTCAAGAACCGCTTTCGAGAAAGCGAGAGCAAGAACTGTGAGAGTCCC 809
764 TGAAGAGCAGTCAAGAAATGCGCTTGAAGAAAGCGAGAGCAAGAAATTCGCGAAATCTG 823
810 TGAAGAGCAGTCAAGAAATGCGCTTGAAGAAAGCGAGAGCAAGAAATTCGCGAAATCTG 869
824 GGAAGTTCCAAAATAGGCAAGTCCGAGACAGACAGACCTCTGTGAAGCAGAGCTGAGG 883
870 GGAAGTTCCAAAATAGGCAAGTCCGAGACAGACAGACCTCTGTGAAGCAGAGCTGAGG 929
884 CTAGTCAGGGAACAGTGTCCAGTCAAGACAGATGACATAGAAATCACTTGGAAAGGA 943
930 CTAGTCAGGGAACAGTGTCCAGTCAAGACAGATGACATAGAAATCACTTGGAAAGGA 989
944 AGGCGAATCAGGGAATAACAGAGAAAAATCCAGGGAAGCGGCAAAACCAAGCTGAGT 1003
990 AGGCGAATCAGGGAATAACAGAGAAAAATCCAGGGAAGCGGCAAAACCAAGCTGAGT 1049
1004 GTGAGGTTTACGACCCCAATGCCCCCTGATCTGCGCCAGCTTCAACAAAGCTGAGT 1063
1050 GTGAGGTTTACGACCCCAATGCCCCCTGATCTGCGCCAGCTTCAACAAAGCTGAGT 1109
1064 TTATGATCTGCTGTGATTCGAGTGAAGAGTGTTCATAGGTGACTGTGGTATTTCTG 1123
1110 TTATGATCTGCTGTGATTCGAGTGAAGAGTGTTCATAGGTGACTGTGGTATTTCTG 1169
1124 AGGCGCGAGGCGGCTCTTGAAGAAAGCGGGAAGACTATCATCTGCGCAAAATTCACCA 1183
1170 AGGCGCGAGGCGGCTCTTGAAGAAAGCGGGAAGACTATCATCTGCGCAAAATTCACCA 1229
1184 TTTTGAAGTGAAGATTAAGACAAACCGGTAGCCCAATGAGCAGAACTCTGGGTGCA 1243
1230 TTTTGAAGTGAAGATTAAGACAAACCGGTAGCCCAATGAGCAGAACTCTGGGTGCA 1289
1244 GATCTGTGGGTGTGATTTGCAAGACAGTCAAGCAATAGGACAGTGAAGACAGAGTCCG 1303
1290 GATCTGTGGGTGTGATTTGCAAGACAGTCAAGCAATAGGACAGTGAAGACAGAGTCCG 1349
1304 GAGAGAACAGGCAATTAAGGGTAGATTAAGAAAGCAAGAAACCCAGCGGCAAGAAAA 1363
1350 GAGAGAACAGGCAATTAAGGGTAGATTAAGAAAGCAAGAAACCCAGCGGCAAGAAAA 1409
1364 AACTCAAGATATTTCAAGCTGTGTGTAAGAGCTCTTGTGTCTTAATGCAATGGAGCTG 1423
1410 AACTCAAGATATTTCAAGCTGTGTGTAAGAGCTCTTGTGTCTTAATGCAATGGAGCTG 1469
1424 GGTGTTCAGTGTGAAGCAAGCTTGACTGTGTGTAATGCAATGGAGCTTCTCAAC 1483
1470 GGTGTTCAGTGTGAAGCAAGCTTGACTGTGTGTAATGCAATGGAGCTTCTCAAC 1529
1484 ACGCAGACGTAACATGAAGATTTCTAAGTTCAAGTAAAGAAACCAAAACCAACAGG 1543
1530 ACGCAGACGTAACATGAAGATTTCTAAGTTCAAGTAAAGAAACCAAAACCAACAGG 1589
1544 AAAAGGTCAAGACGAAGCCGAGAAAGATTCAGTCTTCAAAATGCAAGTTCAGTGGGGA 1603
1590 AAAAGGTCAAGACGAAGCCGAGAAAGATTCAGTCTTCAAAATGCAAGTTCAGTGGGGA 1649
1604 TTTAAATCTCTTCTGTGCAAGAGACTAGCGTCAAGAAAGAGGAAACCCAGTGAAGA 1663
1650 TTTAAATCTCTTCTGTGCAAGAGACTAGCGTCAAGAAAGAGGAAACCCAGTGAAGA 1709

```

Qy	1664	AAGGATGCTGGCTTCCTTCAGAGGTAGAGCTTCCTGGGAAAGAGGACGACCGTGTAGAGCAGCA	1723
Db	1710	AAGGATGCTGGCTTCCTTCAGAGGTAGAGCTTCCTGGGAAAGAGGACGACCGTGTAGAGCAGCA	1769
Qy	1724	CACCATCTGGGGCAAGTACCAACAACCTATTAAGTAAATGTACGTA	1783
Db	1770	CACCATCTGGGGCAAGTACCAACAACCTATTAAGTAAATGTACGTA	1829
Qy	1784	CCACTGCACCTCTGGCCCAACCTATTAAATGTACGTA	1823
Db	1830	CCACTGCACCTCTGGCCCAACCTATTAAATGTACGTA	1869
RESULT 5			
AL732560/c			
LOCUS	AL732560	104663 bp	DNA linear ROD 22-AUG-2002
DEFINITION	Mouse DNA sequence from clone RP24-209P15 on chromosome 2, complete sequence.		
ACCESSION	AL732560		
VERSION	AL732560.11	GI:22474418	
KEYWORDS	HTG.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 104663)		
AUTHORS	Leonardomoralet, D.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humburey@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
COMMENT	On Aug 23, 2002, this sequence version replaced gi:22204552.		

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality ≥ 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-209P15 is from a Male (C57BL/6J) mouse BAC Library VECTOR: pTAKBAC1.

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP24-209P15"
/clone_lib="RPCI-24"

```

ORIGIN

Query Match	43.3%	Score 1241.2;	DB 10;	Length 104663;
Best Local Similarity	99.1%	Pred. No. 0;		
Matches 1269; Conservative	0;	Mismatches 8;	Indels 3;	Gaps 2;

1588 AGTGTTCAGGTGGGATTAATCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAGG 1647

Db	46536	ATTATCCAGGTTGGAGTTTAAATCTCTCTGTCGACCAAGACGTAGCGTCAGAGAAAAGG	462177
Oy	1648	GAAAACCTCAGTGAAGAAAGTGAATGCTGCGCTTCAGAGATGGAACCTTCTGGAGAGAGGCA	1707
Db	46276	GAAAACCTCAGTGAAGAAAGTGAATGCTGCGCTTCAGAGATGGAACCTTCTGGAGAGAGGCA	462178
Oy	1708	GCGTGTAGAGCAGCACAACCATCTCTGGGCAAGTACACAACATCAATCTGTGAAGCCA	1767
Db	46216	GCGTGTAGAGCAGCACAACCATCTCTGGGCAAGTACACAACATCAATCTGTGAAGCCA	46157
Oy	1768	GAGAAGCCAGAGAAAGCCCACTGCACTCTGCGCCACCTTATGATGAATGTAAGTATCAC	1827
Db	46156	GAGAAGCCAGAGAAAGCCCACTGCACTCTGCGCCACCTTATGATGAATGTAAGTATCAC	46097
Oy	1828	CCAAAAGCCTGGCTTCCAGAGGCCCTGCTCCATCAATGAGGAGGCGCTGAGGCGTCTAGG	1887
Db	46096	CCAAAAGCCTGGCTTCCAGAGGCCCTGCTCCATCAATGAGGAGGCGCTGAGGCGTCTAGG	46037
Oy	1888	ACCAGAGTCTCTGGAGTGTCTGTGTGCTGATAGTACAGCAGACCTCACTGCAAGCCAGAAC	1947
Db	46036	ACCAGAGTCTCTGGAGTGTCTGTGTGCTGATAGTACAGCAGACCTCACTGCAAGCCAGAAC	45977
Oy	1948	AGATACCAAGATGCTTGTGACATCCCAAGTGTTCCTGCGCTAAGCTGAGAGCGTCTCTGGG	2007
Db	45976	AGATACCAAGATGCTTGTGACATCCCAAGTGTTCCTGCGCTAAGCTGAGAGCGTCTCTGGG	45917
Oy	2008	TGTTTCTTAAGAGCTGTGTAGAGCGCTCACTGTTGAGAGCAATTTCTATTTCAAGTTTAGG	2067
Db	45916	TGTTTCTTAAGAGCTGTGTAGAGCGCTCACTGTTGAGAGCAATTTCTATTTCAAGTTTAGG	4585
Oy	2068	CCTTGTGATGATTTGAACCTGCTCTCAACAAGTGGGACAATCAGTGCACATCTTTCAAA	2127
Db	45856	CCTTGTGATGATTTGAACCTGCTCTCAACAAGTGGGACAATCAGTGCACATCTTTCAAA	45797
Oy	2128	CTGAGTTCAGACTTGAAGTTCCTCAACAGACAGCCAGCACAACCAATGAGAGTTTCCAC	2187
Db	45796	CTGAGTTCAGACTTGAAGTTCCTCAACAGACAGCCAGCACAACCAATGAGAGTTTCCAC	45737
Oy	2188	GTTACCTATTTTCTTCAACATGTCACCACTGCTCTTCTATAGAAAGTGAATTTTTC	2247
Db	45737	GTTACCTATTTTCTTCAACATGTCACCACTGCTCTTCTATAGAAAGTGAATTTTTC	45687
Oy	2248	ATGAGTTAAGGCAATTTTGAAGAAGACAGTGAACAATTTGTGTCTATTTCAACATCAACA	2307
Db	45679	ATGAGTTAAGGCAATTTTGAAGAAGACAGTGAACAATTTGTGTCTATTTCAACATCAACA	45627
Oy	2308	CACCTACATAAACAAGTGTGAATTTGAATGATCATCTGCTGCACAGTCCCATGTAAGT	2367
Db	45619	CACCTACATAAACAAGTGTGAATTTGAATGATCATCTGCTGCACAGTCCCATGTAAGT	45567
Oy	2368	AATGTGTGTGACCACTACACTACAGACATTTCTCATCAAGTGAAGAAATTTGTGGTGTGA	2427
Db	45559	AATGTGTGTGACCACTACACTACAGACATTTCTCATCAAGTGAAGAAATTTGTGGTGTGA	45507
Oy	2428	GCTTCCAGACCTGTGTGAATGCCAGATTTTCAAAATCAAGTGTATTTAGTTTGAAGTAACTTT	2487
Db	45499	GCTTCCAGACCTGTGTGAATGCCAGATTTTCAAAATCAAGTGTATTTAGTTTGAAGTAACTTT	45447
Oy	2488	TGAAAACCCAAGATTAGATCACTAGATTTGAGGCAATTTCTTGTTCACCTGCTGTACAAGT	2547
Db	45439	TGAAAACCCAAGATTAGATCACTAGATTTGAGGCAATTTCTTGTTCACCTGCTGTACAAGT	45387
Oy	2548	AACACTTCTCTACTTTTGTGAAGTGTGTTCAAGTGAATTTTCACTGTGTGTCTATGAT	2607
Db	45379	AACACTTCTCTACTTTTGTGAAGTGTGTTCAAGTGAATTTTCACTGTGTGTCTATGAT	45327
Oy	2608	AGTGTATTGTTCACATCTGTGTCAACCTCAAGCCCAAGATTAAGGAGGCTCATAGCCTAAG	2667
Db	45319	AGTGTATTGTTCACATCTGTGTCAACCTCAAGCCCAAGATTAAGGAGGCTCATAGCCTAAG	45267
Oy	2668	CTGGCCTTAAATTTTCTGTAAAGTTGGGGGCTCTCTGCCCCCACTGTAAAGAGTGTG	2727
Db	45259	CTGGCCTTAAATTTTCTGTAAAGTTGGGGGCTCTCTGCCCCCACTGTAAAGAGTGTG	45207

Qy	2728	AAACTTACAGGACACACACCTGAACTATACCTGTGGGGGTGTGAATCAAGAGTGTGACACAG	2787
Db	45199	AAACTTACAGGACACACACCTGAACTATACCTGTGGGGGTGTGAATCAAGAGTGTGACACAG	45144
Qy	2788	CAGCTTACACAGCTGAGCTGAGTGTCTTACAGCATGTGTAACAGAACTGCTGATCTTCTGTG	2847
Db	45139	CAGCTTACACAGCTGAGCTGAGTGTCTTACAGCATGTGTAACAGAACTGCTGATCTTCTGTG	45088
Qy	2848	TTTGTTTCTTTGTGTGCTGGG	2867
Db	45079	TTTGTTTCTTTGTGTGCTGGG	45060
RESULT 6			
BD244712		2610 bp	DNA
LOCUS		linear	PAT 17-JUL-2003
DEFINITION		Genes encoding for the human and murine death	
ACCESSION		BD244712	
VERSION		BD244712.1	GI:33054482
KEYWORDS		JP 2002526040-A/1	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS		1 (bases 1 to 2610)	
TITLE		Alonso,C.M., Domingo,D.G., Grandien,A., Leonardo,B. and Martinez,P.	
JOURNAL		Genes encoding for the human and murine death inducer-oblierator-1	
COMMENT		Patent: JP 2002526040-A 1 20-AUG-2002;	
		CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS	
		OS Homo sapiens (human)	
		PN JP 2002526040-A/1	
		PD 20-AUG-2002	
		PR 10-SEP-1999 JP 2000570314	
		PR 10-SEP-1998 SE 9803069-5, 17-SEP-1998 US 60/100873	PI
		CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, PI	
		ESTHER LEONARDO.	
		PI PEDRO MARTINEZ	
		PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC	
		A61P29/00,	
		PC A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC	
		C12N5/10,	
		PC C12P21/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC	
		C12N15/00,	
		PC C12N5/00,A61K37/02	
		CC This gene is referred to in the application as Human Death	CC
		Inducer	
		CC Obliterator Gene 1.	
		CC It has now been named by the Human Gene	
		Nomenclature Committee	
		CC as	
		CC Death-Associated Transcription Factor (DATF-1) FH Key	
		Location/Qualifiers	
		FT source	1..2610
		FT	Location/Qualifiers
		1..2610	/organism='Homo sapiens (human)'.
FEATURES			
source		Location/Qualifiers	
		1..2610	/organism='Homo sapiens'
		/mol_type='genomic DNA'	
		/db_xref='taxon:9606'	
ORIGIN			
		Query Match	35.7%; Score 1023.6; DB 6; Length 2610;
		Best Local Similarity	71.9%; Pred. No. 3.7e-303;
		Matches 1494; Conservative	0; Mismatches 534; Indels 50; Gaps 10;
Qy	1	TCCGTGTGAGCTTCGGAATAGCTCCGGGATCCCGCGCCGGGGAGCTGTGTTAAAGG	60
Db	21	TCCGGGGGCTTCGGGGAATAGCTCCCGAGACCTTGAGAGCCCTGCGAGCTACTTCACGG	80
Qy	61	CAGTCCCACTGCTCCCTTGTGGTGTGCAGACCTCGGAAATCTTCAATGATGAATGCTG	120

Db	81	GAACGCTCTAGATATCTGAATGTTGAAATACGAACGCTGTACTGTGAACGTG	140
Oy	121	ACTGACAAAGTGGGGTAG--GCTTGGCCGTCTGCTTGAAC-----TGCCC	166
Db	141	GCTGACAAAGTGTGTGTGAGCGCTGGCTGTCTGCTTGAACCAAGATTTCTGTGCC	200
Oy	167	CAGGTCTATTAATTTAATGTAAGATTTCCACCAAGTTTCCAAAGTTTCAGTTTGGGA	226
Db	201	AGGGTTTTGTGTATTTTAGAATTT--CAGGAAATGTCTCAAGCTTTCAAGTGTGAG	259
Oy	227	CAGGTATGATGATTAAGGGGCACTGTGACATGAGGAAGCAACCAAGCTATCAACCCA	286
Db	260	CAGGTATGATGACAAAGGGGCACTCCAGCAATGAGAGGCACTTAAGGCATCAACCCA	319
Oy	287	CCAGTAAAGATTCAGGAAAACTCGGGTTTTCGAAGAACCAAGATTGCCAAAGTGAAG	346
Db	320	CCAGCAAAAGTTCAAGAAACAATGGGGTTTTGAGAGACCACTATGCGCAAGGAGAG	379
Oy	347	GTCGAGAGACACGAGAGCGGACCCCACTGACAGCAAC-----ACAGCAGATAAC	400
Db	380	GCGCAGGGGAGACGCGAGGCTGACCCACTGAGACCGCACCCCAACAGCAGCGTGGGCC	439
Oy	401	TCTCCCTGGCCGCGAGTGAACGCGAACCAAAAGTACTGAGAGGGTGAAGAGTTCTTA	460
Db	440	TGTCCTCTGGCCGCGAGTGAAGGCGAACCGCAAGCTGAGCGGTGAAGCATTTCTGA	499
Oy	461	CCAGGTTTCGGCCGCGAGGAGAAAGATATGCGGGTCTCTGAGAGATTCAGTAGAC	520
Db	500	CCATTTGCCCGGCGCGCGGACAGAGAGCAATGCTGTCTCTCTGAGAGATTCGGTAGAC	559
Oy	521	CCACATCTTCCACAGTCACTGATGTGAGACAGCTTCCGAGGGGAGGTTGAAAGCAGTT	580
Db	560	CCAGTCTCGCCCGCACAGACCCCAAGCAGCTCCGAGGGGAGGTTGAAAGCGCTT	619
Oy	581	CTGAGATCAAGATGAGGCGCTGTATCTGTACCTCTTA-----GGGAAAGAACTCTGCGCTT	637
Db	620	CTGAGACCCAGAGGGGCGCCCAAGTCTGTCTTCAAGCTGTGAAGAAAGCAACGCTCTT	679
Oy	638	CTGAAAAGGCAAAAGAGAGTGAAGAGAAAGAACACTCTTGAAGTGAAGTAGCGCC	697
Db	680	CTGAAAAGGTGAAGAGAGGGATGACCAAGATGACACTCCGATAGTGAACAGCGATGCGC	739
Oy	698	TTACGTTGAAGGAATTCAAGACCGCTTCGGAAGAAAGCGAGCAAGAACCTGTGAGA	757
Db	740	TGACCTTGAAGAGCTTCAAGATGCGCTTCGCAAGAAAGCGGAACAGAGCCCACTGAGA	799
Oy	758	GATCCCTGAGAGGCGAGTCAAGATGCGCTGAGAGAAAGAGCGCAGAGAGAGATTTCTCGC	817
Db	800	GGCCCTTGAAAGGGATGCCAGATGCGCTGCGGAAGAAAGCGCCGGAAGAGAGGTCCCGCG	859
Oy	818	AAACTGGGAGTGTCCAAATAGGCAAGTGCAGAGCGAGACAGACCTCTGTATGAGAGAGC	877
Db	860	AGACTGTGGGCTCCGAGGCCAGTGAACATGTGTGAGGGCGTCTGCGCAATTAAGCAGAGC	919
Oy	878	CTGAGGCTAGTCAAGGACCAAGTGTCCCAAGTCAAGACAGATGACATGAAAAATCAATTGG	937
Db	920	CCGAGAACGATCAAGGGGGTGTGTCTCCAGGCTGAGAAAGATGACAGAGAGTAAAGTTGG	979
Oy	938	AAGGAAAGGCGACTCAGGGAAATACAGAGGAAACCCCAAGGAAAGCGGCAACCAAGC	997
Db	980	AGGAAAGGCGGCTCAGGACATCAAAAGATGAGAGCGCTGAGAGCTTGTGGCGCAGCCGAAGC	1039
Oy	998	CTGAGTGTGAGGTTTGAAGCCCAATGCGCTGTATCTGATCTGTGCGCGCAGCTCAACA	1055
Db	1040	CTGATGTGAGGGTTACAGACCCCAAGCGCTGTATGTGATTTGTGCGCGCAGCTCAACA	1099
Oy	1058	ACAGGTTATGATCTGCTGTGATCGGTGTGAGAGTGGTTCATGTGTGATCTGTGTGGTA	1117
Db	1100	ACAGGTTATATATTGCTGTGACCGCTGTGAAGATGGTTTCAATGGGCAATGTGTGTGGCA	1159
Oy	1118	TTTCTGAGGCGCGAGGCGGCTCTGTGAAGAAAGCGGGAAAGATCAATCTGCCCAATT	1177
Db	1160	TTTCTGAGGCTGAGGGAGGCTTTTGAAGAAAGATGGGGAAGATATATCTGCCCAACT	1219

QY 1178 GCACATTTTGCAGTGAAGATGAGAAACGGTAGGCCCAATGAGCAGACTG 1237
 DB 1220 GCACATTTTGCAGTGAAGATGAGAAACGGTAGGCCCAATGAGCAGACTG 1279
 QY 1238 GGTGAGATCTGTGGGTCTGATGAGCAGACTGAGCAGATGAGGACATGAGCAG 1297
 DB 1280 AATGAGACCTGAGATGATGAGCAGACTGAGCAGATGAGGACATGAGCAG 1339
 QY 1298 AGTCGAGAGAGACAGAGGATGAGGATGAGGAGCAGCAGCAGCAGCAG 1357
 DB 1340 AGTCGAGAGAGACAGAGGATGAGGATGAGGAGCAGCAGCAGCAGCAG 1399
 QY 1358 AGAAAAAATCTGAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAG 1417
 DB 1400 AGAAAAAATCTGAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAG 1459
 QY 1418 GCGCTGGGTGTTCCAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAG 1477
 DB 1460 GCGCTGGGTGTTCCAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAG 1519
 QY 1478 TCAAAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1537
 DB 1520 TCAAAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1579
 QY 1538 CCAAGGAAAGGTCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1597
 DB 1580 CCAAGGAAAGGTCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1639
 QY 1598 TGGGATTAATCTCTCTGTCAGAGAGCAGCAGCAGCAGCAGCAG 1657
 DB 1640 TGGGATTAATCTCTCTGTCAGAGAGCAGCAGCAGCAGCAGCAG 1699
 QY 1658 TGAAGAA---AGTATGCTGCTTCAAGAGCAGCAGCAGCAGCAG 1714
 DB 1700 TGAAGAAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1759
 QY 1715 AGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1774
 DB 1760 AGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1819
 QY 1775 CAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1834
 DB 1820 CAG---CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1870
 QY 1835 CTGGCTTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1889
 DB 1871 TTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1930
 QY 1880 CAG-----AGTCCTGGGTGCTCTGAGCAGCAGCAGCAGCAG 1940
 DB 1931 TGGGATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1990
 QY 1941 CAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2000
 DB 1991 CAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2050
 QY 2001 CTCTGGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2038
 DB 2051 TCT 2088

RESULT 7
 AK023369 2610 bp DNA linear PAT 15-SEP-2000
 LOCUS AK023369 Sequence 1 from Patent WO0015787.
 DEFINITION AK023369
 ACCESSION AK023369
 VERSION AK023369.1 GI:10183781
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1
 AUTHORS Leonardo, R., Martinez, P., Alonso, C.M., Domingo, D.G. and Grandien, A.
 TITLE Genes encoding for the human and murine death in
 JOURNAL ducer-obliterator-1
 PATENT: WO 0015787-A 1 23-MAR-2000;
 LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR
 INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO
 CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALF (GB)
 FEATURES
 source
 1.2610
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="This gene is referred to in the application as
 Human Death Inducer Obliterator Gene 1. It has now been
 named by the Human Gene Nomenclature Committee as
 Death-Associated Transcription Factor (DNF-1)"
 ORIGIN
 Query Match 35.7%; Score 1023.6; DB 6; Length 2610;
 Best Local Similarity 71.9%; Pred. No. 3.7e-303;
 Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;
 QY 1 TCCGTGAGTCTCTGAGAAATGCTGCGGATCCCGCGCGGAGCTTTAGAG 60
 DB 21 TCCGCGCGCTTCCGAGAAATGCTGCGGATCCCGCGCGGAGCTTTAGAG 80
 QY 61 CAGTCCCACTGCTCCCTTGTGTTGCGAGCTCCGGAATCTTCTCATGAGAACTG 120
 DB 81 GAAACAGCTCTGATTAATCTGAGTTGTTGAAATACAGAGCTTTCTGTAACAG 140
 QY 121 ACTGACACAGTGGGAGTGAAG-CTTGGCGCTCTGCTTACG-----TGGCC 166
 DB 141 GCTGACACAGTGTGTTGTTGAGAGCTTGGCTGCTGCTTGAACAGAGTTTCTG 200
 QY 167 CAGTCTTAATTTTATGATGAGATTTCAAGCAGAGTTTCAAGTTTCAAGTTT 226
 DB 201 AAGGTTTGTGTTGATTAATGAGATTT-CAGGAAAGTGTCAAGTTTCAAGTTT 259
 QY 227 CAGGTTATGATTAATGAGGAGCAGTGAAGATGAGGAGCAGCAGAGCTTATCA 286
 DB 260 CAGGTTATGAGCAGCAGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 319
 QY 287 CAGTAAAGAGTTTCAAGAAACCTGCGGTTTTCAGAAACAGATTTGCAAG 346
 DB 320 CAGGAAAGAGTTTCAAGAAACATGCGGTTTTCAGAAACAGATTTGCAAG 379
 QY 347 GTGACAGAGCAGCAGAGCGGAGCAGCAGCAGCAGCAGCAGCAGCAG 400
 DB 380 GCGCAGGAGCAGCAGAGGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 439
 QY 401 TCTCCCTGCGCGCAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 460
 DB 440 TCTCCCTGCGCGCAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 499
 QY 461 CCAAGGTTGCGCGCAGGAGGAGAAAGATGTCGCGGTCTCTGAGATTTCAAG 520
 DB 500 CCAATGCGCGCGCGCGCGCAGGAGGAGATGCTCTCTGAGGATTTCAAG 559
 QY 521 CCAATTTTCCAGATCATGATGATGAGAGCAGCTTCCGAGGAGGAGGAGGAG 580
 DB 560 CCAAGTCTGCGCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 619
 QY 581 CTGAGTCAAGAGTGGCTCTGATCTGATCTCTTA---GGAAAGACATCTGCT 637
 DB 620 CTGAGTCAAGAGGCGCGCGCGCAGTCTCTTCCAGCTGATGAGAGCAGCAG 679
 QY 638 CTGAAAGAGCAGAAAGAGTGAAGAGAGAGAGCAGCTTGAAGATGAGTGG 697
 DB 680 CTGAAAGAGTGAAGAGAGGAGTGAAGAGAGAGCAGCTTGAAGATGAGTGG 739
 QY 698 TTAAGTGAAGAGCTTCAAGAGCAGCTTCCGAGAGAGGAGAGAGAGAGAG 757

Db 740 TGAACCTGAAAGAGCTTCAGAAATCCCTTCGACAGAGACCGGAGACAGAGGCCCACTGAGA 799
 Qy 758 GGTCCCTGAGAGGGAGTGAAGATGCGCTGAGAGAAAGCGAGAGAGAAATTCGCG 817
 Db 800 GGGCCCTGAAAGGGATCCAGAGTCCCTGCGAGAGAAAGCGGAGAGAGGTCCCGCG 859
 Qy 818 AACTGAGAGTGTCCAAATAGGACATGCGAGACAGACAGACCTCTCTGTAAAGCAGAGC 877
 Db 860 AGACTGTGGGCTCCGAGGCGCAGTGA CACTGTGAGAGGGGTCTGCGCCAGTAAGCAGAGC 919
 Qy 878 CTGAGGCTAGTACAGGACCACTGTCTCCAGTCAAGACAGATTAATTAATAATCACTTG 937
 Db 920 CCGAGAACGATCAGGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGATTAAGTGG 979
 Qy 938 AAGGAGAGGCGACTCAGGAAATACAGAGAAACCCAGAGGAGCGGCAACCAAGC 997
 Db 980 AGGAAAGGCGGCTCAGACATCAAAAGATGAGAGGCTGAGAGCTTGGGCGACCGAGC 1039
 Qy 998 CTGAGTGTAGGTTTACAGACCCCAATGCCCTGTACTGCAATGCGCGCAGCCTCAACA 1057
 Db 1040 CTGAATGTAGAGGTTAGACCCCAAGCGCTGTATTTGATTTGCCGACGCTCAACA 1099
 Qy 1058 ACAGGTTTATGATCTGTGTGATGAGAGTGTGAGAGAGTTCATATGTTGATGTTGGGTA 1117
 Db 1100 ACAGGTTTATGATTTGTGTGATGACCGCTGTGAAGATGTTTCAATGCGATGTGTGGCA 1159
 Qy 1118 TTTTCTGAGGCGCGGCGGCTCCCTGGAAGGAAAGGAGGAGAACTCAATCTGCGCAAT 1177
 Db 1160 TTTCTGAGGCTGAGAGGCGCTTTTGGAAAGGAAAGGAAAGCTATATCTGCGCAACT 1219
 Qy 1178 GCACCATTTTCAAGTGCAGATGAGACAAACGATGAGCGCCCAATGAGCAGAGCTCG 1237
 Db 1220 GCACCATTTTCAAGTGCAGATGAGACAAACGATGAGCGCCCAATGAGCAGAGCTCG 1279
 Qy 1238 GGTGCAAGTCTGTGGGTCTGATGAGCAACAAGCTGCAACAAGCATGAGAGAGCAGA 1297
 Db 1280 AATGAGAGCTGAGATGCTGATGAGCAGCCGATTTGATCAAGTATGAGAACTAATGAGCAGA 1339
 Qy 1298 AGTCCGGAAGACCAAGGAGCAATTAAGGATGAGAAAGGAGCAAAACCCAGCGCA 1357
 Db 1340 AGTCTAGGAGAGCAAGGAGCAATTAAGGATGAGAAAGGAGCAAAACCCAGCGCA 1399
 Qy 1358 AGAAGAACTCAAGATATTCAGCTGTCTGTGAGAGGCTCTCTGTCTCTTAATGATTTG 1417
 Db 1400 AGAAGAACTCAAGATATTCAGCTGTCTGTGAGAGGCTCTCTGTCTCTTAATGATTTG 1459
 Qy 1418 GCCCTGGGTGTTCCAGTGTAGACAGCTGACTGTGTGTATTTGAGTAATGACTGCATTC 1477
 Db 1460 GCCCGGGGTGCTGTCACTGGGCGCAGCCGCACTCGGTATCTGCAGTAATGACTGTATCC 1519
 Qy 1478 TCAAAACCGGAGGCTCAATGAGATTTCTAAGTTCCAGGTAAGAAACAAACAAAC 1537
 Db 1520 TCAAAACCGGAGGCTCAATGAGATTTCTAAGTTCCAGGTAAGAAACAAACAAAC 1579
 Qy 1538 CCAAGAAAGGCTCAAGAGCAAGCAAGAAAGTTCACTCTTCCAAATGCAAGTGTTCAG 1597
 Db 1580 CTAAGAAAGGCTCAAGAGCAAGCAAGAAAGCCAGTCTTCCGAAATGCGGTGCTCAG 1639
 Qy 1598 TGGGGAATTAATCTCTTCTGTGCAAGAGCTAAGCTCAGAGAAAGGAGAAACCCAG 1657
 Db 1640 CAGGTAATTAATCTCTTCTGTGCAAGAGCAAGCTCAGAGAAAGGAGAAACCCAG 1699
 Qy 1658 TGAAGAA---AGTATGTGCTTCCAGAGTGAACATTTGAGGAAAGGAGGAGCTGTG 1714
 Db 1700 TGAAGAAAGGAGTGTGTCTCTGTGCGAGAGTGAAGCACTCGGAAAGGAGAGCTGTG 1759
 Qy 1715 AGAGAGACACACATCTGTGGCAAGTGAACAACATCAATCTGTGAGAGCCAGAGAAC 1774
 Db 1760 AGAGAGACACCCCGTGTGGGAGAGATCAATTAATCAATCAGATTAAGCCAGAAAGA 1819
 Qy 1775 CAGAGAGCCCACTGCTGTGCCCACTTATGTAATGTAATGTAATGTAATGTAATGTAATG 1834
 Db 1820 CTG-----CTGCTCTCTGCGCGTCACTGTGTATTAATGTAATGTAATGTAATGTAATG 1870

Qy 1835 CTGCTTCCAGAGCCCTCTCCATCATCT-----GGTGGCTGCTGTGGGCGCTGTACAGAC 1889
 Db 1871 TTGGCTCTCTGAGACCCCTCCGCTTCTTTCTGATAGCCATCCCTGGGCGCTGTCCAGAG 1930
 Qy 1890 CAG-----AGTCTGGGTGTTCTGTGTCTGTATGATGATGATGATGATGATGATGATG 1940
 Db 1931 TGGGAGTTCACACTTGTGTGTATGAGCTGATCAGACACCGGCTGACCACTCAGCGGAG 1990
 Qy 1941 CAGAGCAGATATACAAAGATCCCTGTGACCCAGGTTCTCTGCTAGGCTGTGAGCCT 2000
 Db 1991 CAGAGCCCAATGTCAGATGCTCTCTGCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGT 2050
 Qy 2001 CTGTGGGTGTTCTTAAGAGCTGTGTAGGCTCATGT 2038
 Db 2051 TCTGTCACTGTTTCCAAAGCTGTAAACCTCATCTGT 2088
 RESULT 8
 BC004237 2767 bp mRNA linear PRI 12-NOV-2003
 LOCUS
 DEFINITION
 Homo sapiens death associated transcription factor 1, transcript variant 1, mRNA (cDNA clone MGC:896 IMAGS:2959493), complete cds.
 ACCESSION
 BC004237 GI:33872756
 VERSION
 BC004237.2 GI:33872756
 KEYWORDS
 MGC.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Bularctica; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 2767)
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh P., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Brownstein M.J., Utsid T.B., Toshiyuki S., Scheetz T.E., Brownstein M.J., Utsid T.B., Toshiyuki S., Carinici P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Mizny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green B.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywinski M.I., Skalska U., Smalins D.B., Scherch A., Schein J.B., Jones S.J. and Matre M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 JOURNAL
 MEDLINE
 PUBMED
 12477932
 2 (bases 1 to 2767)
 REFERENCE
 Strausberg R.
 Direct Submission
 Submitted (01-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:13278980.
 CONTACT: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMLT)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>
 CONTACT: amadan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: <http://image.llnl.gov>
 Series: IPAL Plate: 2 Row: 0 Column: 16
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18375620.

FEATURES

SOURCE

1. 2767
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:896 IMAGE:2959493"
 /tissue_type="Colon, adenocarcinoma"
 /clone_lib="NIH MGC_15"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 1. 2767
 /gene="DATP1"
 /note="synonyms: D101, FLJ11265, KIAA0333"
 /db_xref="locusid:11083"
 /db_xref="MIM:604140"
 282. 1970
 /codon_start=1
 /product="death associated transcription factor 1, isoform a"
 /protein_id="AAH04237.1"
 /db_xref="gi:13278981"
 /db_xref="locusid:11083"
 /translation="MDDKDGSPENBAPAIKPTKFRKTMGRFRTTIKREAGDAE
 ADLEPPPOQIGILSRSGRPSRQSAIRVAPASISVGVGGDSDSGILT
 PATDAETBSBSVBSASSTRSGPSAIVKREKREBPATVBSADTVETVLPKOR
 LNEIDNRIRKREDEPTERPLKIOSIRLAKRREBPATVBSADTVETVLPKOR
 PENQGVVSQAKGDKRESLREKAADI KDEPGLGRPECEYDPAALCTICOP
 HNNRPMICCDRCBWFHGVGISEARGLERNGSDYICPNCTIIQVDETSBAD
 OOEAKRPGDADGTDCTSIGTEKSGSEDOGIKRIERKANSCKKXIPQVIBAP
 GASKICPGCCGAAVDVYVSGNDCLIKHAATMKELSGEKORPEKEMKMEPE
 SLPRKGAAGIKTISVHRPRAPKEKETTAKAVVVPASBAIKREKAGESSPSMAD
 HNYANVFKRTAPSPSLYKCKTHLGVGLDPSRSFWAIPTWACPGALVLC"
 1092. 1241
 /note="PHD; Region: PHD-finger. PHD folds into an
 interleaved type of Zn-finger chelating 2 Zn ions in a
 similar manner to that of the RING and FYVE domains"
 /db_xref="cdm:pfam00628"

CDS

gene

misc_feature

1092. 1241
 /note="PHD; Region: PHD-finger. PHD folds into an
 interleaved type of Zn-finger chelating 2 Zn ions in a
 similar manner to that of the RING and FYVE domains"
 /db_xref="cdm:pfam00628"

ORIGIN

Query Match 35.4%; Score 1016.2; DB 9; Length 2767;
 Best Local Similarity 73.5%; Pred. No. 7, 2e-301;
 Matches 1414; Conservative 0; Mismatches 474; Indels 36; Gaps 8;

141 GCTTGCCGCTCTGCTTGAACCTGAGCCCAAGTCTATTAATTATTAATGATTTCCAGCCAA 200
 |||||
 192 GCTTGAGACCCAGAGGTTCTGCTGCGCAGGGTCTTGTGTGTTAGATTT-CAGGGAA 250
 |||||
 201 AGTTTCAAGCTTCAGTTTGGGACAGGTATGATGATTAAGGGCACTGAGCAATGA 260
 |||||
 251 AAGTGTCCAAAGTTTCAGTTGAGACAGGTATGACGACAAAGGAGCACTCGACATGA 310
 |||||
 261 GGAAACACCCAAAGGCTATCAACCCACAGTAAGAGTTCAAGAAAACTGGGGTTTTG 320
 |||||
 311 GGAGGACCTTAAGGCATCAAAACCCACGAAAGAGTTCAAGAAAACTGGGGTTTTG 370
 |||||
 321 AAGAACCAAGATTGCAAACTGAGAGGTGACAGAGACAGGAGCGGACCCAGAGAGA 380
 |||||
 371 AAGGACCACTATTCGCAAGAGAGGCGCAGAGGACGCGAGAGCTGACCCACTGAGTCC 430
 |||||
 381 GCAACCC-----ACAGACGATTAACCTCTCCCTGCGCGCAGTGAAGGCAACCAAAAG 434
 |||||
 431 GCCACCCCAACAGACAGCACTGGGCTGTCTCTGCGCGCAGTGAAGGCAACCCAAAGG 490
 |||||
 435 TACTGAGAGGTGAAGAGTTTCTTAACAAGTTGGCGCCCGAAGGAAAAAGATGTGCC 494
 |||||
 491 CACTGAGCGCGTGAAGAGTCTCTGACCAATTCGCGCGCGCGCGAGAGAGGAGTACC 550
 |||||

495 GGTGTCCTGAGAGATTCAGTAGAGCCCACTTCCACAGTCACTGATGTGAGACAC 554
 |||||
 551 TGTCTCTCTGAGAGATTCGTGTGAGCCACAGCTCTGCCCCGACAGACGCCGAGACAC 610
 |||||
 555 TTCCGAGGGAGCGTTGAAAGAGGTTCTGAGATCAGAAAGTCCCTGTATCTGACTCTT 614
 |||||
 611 CTCCGAGGGAGCGTGAAGAGCGCTTGTGAGACACAGAGCGGCCCTCACTGTCTCCAC 670
 |||||
 615 A---GGAAAGAACATCTCTGCTCTTTTGAAGAAAGGCAAGAGGTGAAGAGAGAA 671
 |||||
 671 AGCTGTGAAGAAACACACAGCTCTTGTGAAGAAAGGTGAAGAGAGAGATGACCAATGA 730
 |||||
 672 CACCTCTGACAGTGAACAGTATGAGCTTACGTTGAAGAACTTCAAGACCGCTTCGAG 731
 |||||
 731 CACCTCCATGATGATGACAGGATGAGCTTGAACCTTGAAGAGCTTCAAGAAATGCTTCGAG 790
 |||||
 732 AAAGCGAGCAAGAACCTGTGAGAGGTCTCTGAAGGCACTCAGAAATGCGCTTGAAGAA 791
 |||||
 791 GAAGCGGAAACAGAGGCCCACTGAGAGGCCCTGAAGGGAATCCAGATGCGCTGCGAA 850
 |||||
 792 GAAGCGCAGAGAGAAAGATTTCTGCGGAAACGTGGAGTGTCCAAATGAGAGTCCGAGCA 851
 |||||
 851 GAAGCGCGGAGAGAGGTCCCGCAGACGTGTGGCTTCGAGGCCAGTGAACCTGTGGA 910
 |||||
 852 GGACAGACCTCTCTGTAAGCAGAGAGCTGAGGCTAGTCAGGAGCCAGTGTCCAGTCA 911
 |||||
 911 GGCGCTCTGCGCCAGTGAAGAGAGCCCGAAGAACATGCGGGGTTGTCTCCAGGCTGG 970
 |||||
 912 GACAGATGACATGAAGAAATCAGTTGGAAGAGAGGCGACTCAGGGAATACAGAGAAA 971
 |||||
 971 GAAAGATGACAGAGAGATGATGAGAGGAAAGCGGCTCAGAGATCAAGATAGAGA 1030
 |||||
 972 CCCGAGGAGAGCGGCAAAACCAAGCTGAGGTGAGGTTCAGACCCCAATGCCCTGTA 1031
 |||||
 1031 GCTGGAAGCTTGAGCCGACCAAGCTGAAGTGAAGGTTCAGACCCCAACCCCTGTA 1090
 |||||
 1032 CTGACATCTGCCCGCAGCTCAACAACAGTTATGATCTGCTGATCGGTGAGGA 1091
 |||||
 1091 TTGCATTTGCCCGCAGCTCAACAACAGTTATGATTTGCTGTGACCGCTGTGAAGA 1150
 |||||
 1092 GTGTTCCATGTGACCTGTGTGGATTTTGTAGAGCCCGAGGCGGCTCTGGAAGGA 1151
 |||||
 1151 ATGTTTCAATGCGATTGTGTGGCAATTTCTAGAGCTCAGAGGAGGCTTTTGAAGGA 1210
 |||||
 1152 CGGGAAGACATCACTGCGCCCAATTTGACACATTTTGAAGTGCAGAGTGAACAAACG 1211
 |||||
 1211 TGGGAAAGCTATATCTGCGCAACCTGACCAATTTGCAAGTGCAGAGTGAAGCTATTC 1270
 |||||
 1271 AAGAAAGGAGATCAGCAGAGAGCTTAATGAGACCTGAGATGCTGATGGACCGATTG 1330
 |||||
 1272 CAAGACATAGGAGACAGTAGACAGAAAGTCCGAGAAAGCCAGGGCATTAAGGGTGAAT 1331
 |||||
 1331 TACAAGTATAGAAACAATAGACAGAGAGCTAGCGAAAGCAAGGATTAAGGGTGAAT 1390
 |||||
 1332 TGAAGAGGAGCAAAACCCAGAGGCGCAAGAAATAAATCAAGATATTCAGCTGTGTA 1391
 |||||
 1391 TGAAGAAAGCTGAATAATCAAGTGGCAAGAGAACTCAAGATCTTCAGCTGTGATGA 1450
 |||||
 1392 GGTCTGTGTCTCTTAATGATGAGTGGCCCTGGGTGTTCAAGTGTAGCAACGCTGACTC 1451
 |||||
 1451 GGGCTGTGTCTCTTAATGATGAGTGGCCCGGGTCTGTCACTGGCCAGCCGACTTC 1510
 |||||
 1452 TGTGATATGAGTATGACTGATTTCTCAACCGCAGAGCTTACATGAGATTTCTTAAG 1511
 |||||
 1511 GGTGATATGAGTATGACTGATTTCTCAACCGCAGAGCTTACATGAGATTTCTTAAG 1570
 |||||
 1512 TTCAAGTAAAGAAACAAACCAAGGAAAGGCTCAAGACGAGCAAGCAAGAAAGTT 1571
 |||||
 1571 CTGAGTAAAGAAACAGAACCCAAAGCTTAAGAAAGATGAAGATGAACCGAGAGAGCC 1630
 |||||
 1572 CAGTCTTCCAAATGACAGTGTCAAGTGGGATTAATCTCTTCTGTGCAACAAGAGCT 1631
 |||||

```

AUTHORS Strassberg, R.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT Contact: MGC help desk
Email: cga@dcsc.fda.nih.gov
Tissue Procurement: ATCC
DNA Library Preparation: Rubin Laboratory
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada
Info@dcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywnski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McAleavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES
Source
location/Qualifiers
1. .2725
/mol_type="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:23216 IMAGE:4901057"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_id="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: POT87"
1. .2725
/gene="DANF1"
/note="synonyms: D101, FLJ11265, KIA0333"
/db_xref="LOCUSID:11083"
/db_xref="MIM:604140"
238. .1926
/codon_start=1
/product="death associated transcription factor 1, isoform a"
/protein_id="AAH14489.1"
/db_xref="GI:15680267"
/db_xref="LOCUSID:11083"
/translation="MDKQDPSVNSBPALIKPTSKPRKTKWGFRRITIAKRGAGDAEADLPKPPPOOGLSLRSGRQPKRTERVQEOALTARRRGGRSPVSLDSGPTSCPATDARTASRGSVRSASSTRSGQSASTAKRBPASKEKGGDHDHDSDSIDGLTLEKLNRLRKREKQEPTEPRFKIGIKQSLRKRRBGGPAETVGSBASDTVGCLPSKQBPENRQVSVQAGDDRRSRKLGKGAQDIDKEEGDGLRPPRCBGVDPNALYICRQDPHNRPVICDRRCRWMFHGDCVIGSEARGLIENGSDYICPNCTILQVDETHSICRQOQAKRRPGADGTDGTSIGTIRKSSSDGIGKRIEKANPGKKGLFQVPIEAPRGAQKRGCCCHCAOPDSVCSNDILKHAATVMKFLSSGKQKQPKRGMKPKRKSLEPKCAQAGIKTSSVHKRPRAPRKETTYKCAVVPARSRALAKREACSTSPWADHNNNAVPEKRTAPSPSLIKKYNHILGVLGIDPSRSFWIAPNACPLGVALLC"
1048. .1197
/note="PHD; Region: PHD-finger. PHD folds into an interleaved type of Zn-finger chelating 2 Zn ions in a similar manner to that of the RING and FYVE domains"
/db_xref="CCD:pfam00628"

Query Match 35.4%; Score 1013.6; DB 9; Length 2725;
Best Local Similarity 73.5%; Pred. No. 4,5e-300;
Matches 1414; Conservative 0; Mismatches 474; Indels 36; Gaps 8;
ORIGIN

```

QY	141	GCTTGGCCGCTCTGCTTGA	CTGAGCCCGCCAGGCTCTAATTTTATGTAGGATTTCCAGCCA	200
Db	148	GCTTGGACCCAGAGGTTTCGTC	TCAGAGGTTTGTGTGATTTTAGATTT - CAGGAAA	205
QY	201	AGTTTCCAAAGCTTTCAGTTT	TGGGACAGGATGTGATGTATTAAGGGACCTGAGCAATGA	260
Db	207	AAGTTCGAAGCTTTCAGTTGT	TGGAGCAGGTATGACACAAAGCCAGCCAGATGA	266
QY	261	GGAAGCACCCAGAGCTATCA	AAACCCACAGTAAGGATTCAGAAAACCTGGGGTTTTG	320
Db	267	GGAAGGACCTTAAGGCAATCA	AAACCAACAGAGAAAGATTCAGAAAACAATGGGGTTTTG	326
QY	321	AAGAACCAACGATTGCCAA	ACGTGAGGGTCAGAGACACAGAGCCGACCCCAATGAGCA	380
Db	327	AAAGACCACTATCGCCAA	AGCAGAGGGGACAGGGACCCGAGGCTGACCCACTTGAGGCC	386
QY	381	GGAACCC-----ACGAG	CACTAACTCTCCCTCTGGCCGCGAGTGGACGGCAACAAACG	434
Db	387	GCAACCCCAACAGACAGAC	AGCTGGGCTGTCCCTCTGGGCGCAATGGAGGCAAGCCCAAGC	446
QY	435	TACTGAGAGGGGTGAGAG	ATTCTTACACGGTTTCGGCCCGAGGAGAAAAGAAATGTGTC	494
Db	447	CACCTAGAGCCGTGAGAG	CAATTCACATTCGCGGGCCGCGCAGAGAGAGATGTC	506
QY	495	GGTGTCCCTGAGAGATTCA	AGTCCAGCCCACTTTCACAGTCACTGATGTGTGAGACAGC	554
Db	507	TGTCTCCCTGAGAGATTCT	GTGTGAGACCCACGCTCTGCCCCGACAGACGCGCGAGACAGC	566
QY	555	TTCCGAGGGGAGCGTTTGA	AAAGAGTTCTGAGATTCGAAGTGGCCCTGTATCTGACTCCTT	614
Db	567	CTCCAGGGCAACGTGTGA	AAAGCGCTTCTGAGACACAAAGCGGCCCCCACTGTGCTTCAAC	626
QY	615	A---GGGAAAGAACATTC	CTGCTCTTCTGAAAAGCAAAAGAGGTGAAAGGAGAAAGA	671
Db	627	AGCTGTGAAGGAAACAC	ACAGCCCTCTTCTGAAAAGGTGAAGAGGAGATGACACAGATGA	686
QY	672	CACCTCTGACAGTGA	CAGTGAATGACCTTACGTTGAAGAACTTCAGAAACGCTTCGAG	731
Db	687	CACCTCCGATAGTGA	CAGGATGAGCTGACCTTGAAGAAAGCTTCAGAAATGCGCTTCGAG	746
QY	732	AAAGGAGAGCAAGAACT	GTGGAAGGTCCTCTGAGAGGCAAGTCAAGATGCGCTGAGAA	791
Db	747	GAAAGCGGAAACAGAG	ACCACTGAAGAGCCCTGAAAGGAGTCCAGAGTGCCTGCGGAA	806
QY	792	GAAAGCAGAGAGAGAA	GATTTCTGCCGAACTTGAGGTGTCCAAATGAGCAGTGCAGACA	851
Db	807	GAAAGCCCGGAGAGAG	GGTCCGCCGAGACTGTGGGCTTCCAGAGCCAGTGAACACTGTGGA	866
QY	852	GGAACAACCTCTCTGT	TAAGCAGAGACCTGAGGTGTAGTCAAGGACCAAGTGTCCCACTGA	911
Db	867	GGGCGTCTGCCCAG	TAAAGCAGAACCCGAGAACATGAGGGGGTGTGTCCCAAGCTGG	926
QY	912	GACAGATGACATAGAA	AAATCAAGTTGGAAAGGAAAGCGACTCAGGGAATATCAGAGGAAA	971
Db	927	GAAAGATGACAGAGAG	AGTAAATTGAGAGGAAAGCGCTCAGGACATCAAAAGATGAGGA	986
QY	972	CCGACAGGAAAGGGG	CAAAACAAAGCTGAGTGTAGAGTTTACAGCCCTCAATGCGCTGTA	1031
Db	987	GCTTGAAGACTTGGG	CCAGCCAGACCTGAAATGTAGGGTTTACAGCCCAACGCGCTGT	1046
QY	1032	CTGCACTTGC	CCGACAGCCCTCAACAACAAGTTTATGTATCTGTGTGAGGA	1091
Db	1047	TTGCAATTG	CCGACAGCCCTCAACAACAAGTTTATGTATCTGTGACCGCTGTGAGGA	1106
QY	1092	GTGTGTCCATGTG	ACTGTGTGGGTATTTCTAGAGCCGAGAGGGCGGCTCTTGGAAGGA	1151
Db	1107	ATGGTTTCA	TGCGATGTGTGTGGGCAATTTCTGAGGCTGAGAGGAGGCTTTTGGAAAGGA	1166
QY	1152	CGGGAAGACTACATCT	GCACAAATTTGACCAATTTGCAAGTGTGACAGATGAGACAAACG	1211
Db	1167	TGGGAAAGACTATAT	CTTCCCAAACTGCAACCAATTTCTGCAAGTGTGAGATGAGACTATTC	1226
QY	1212	TAGCCCAACAAATG	ACAGAGACTCTGTGGGTGAGATCTGTGGGTGCTGATGCAACAAGCTG	1271

Db	1227	AGAAACGGCAGATACAGCAGAGAGCTAAATGAGAGCTGGAGATGCTGATGGCAACGANTTG	1286
Qy	1272	CACAGCATNAGGACAGTAGAGCAGAAAGTCCGAGAAAGACGAGGCATTAAGGTTAGAT	1331
Db	1287	TACAAAGTATAGGACAAATAGAGCAGAAAGTCTAGCGAAGACCAAGGGATTAAGGGTAGAAT	1346
Qy	1332	TGAGAGGCGACGAAACCCGACGGCAGAAAGAAAATCTAAGTATTCCAGCCTTGTCGTA	1391
Db	1347	TGAGAAAGCTGCAAAATCCAAAGTGGCAGAAAGAACTAAGATCTTCCAGCCTTGATAGA	1406
Qy	1392	GGCTCTGGTGTCTCTCTAAATGATCTTGGACCTCGGGTGTTCAGTGTAGACAGCCTGACTC	1451
Db	1407	GGCGCTGGATGCTCTCAAAATGTATTTGGCCCCGGGTGCTGTCAAGTGGCGAGCCGACTC	1466
Qy	1452	TGTGTATTGCAGTATGACTGCAATTCCTCAACACGCGACAGACTACATGAGATTCTTAAG	1511
Db	1467	GGTGATCTGCAGTATGACTGTATTCCTCAACACGCGACAGCAGAAATGAAGTTCTTAAG	1526
Qy	1512	TTTCAGGTAAAGAACAAAGAAACCAAAACCCAGAGAAAGGTCAAGACGAAAGCAAGAAAGTT	1571
Db	1527	CTCAGGTAAAGAACAAAGCCAAAGCCCTTAAAGAAAGATGAAGATGAAGCCAGAGAAAGCC	1586
Qy	1572	CAGTCTTCCAAATAGCAGTGTTCAGGTGGGAGTTAAATCTCTCTGTCACAGAGACT	1631
Db	1587	CAGTCTTCCAAATAGCGGTGCTCAGGCGAGGATTAATTAATCTCTCTGTCACAGAGACC	1646
Qy	1632	AGCCTCAGAGAAAGGAGAAACCCAGTGAAGAA--AGTATGCTGGCTTCAGAGATGA	1688
Db	1647	AGCTTCAGAAAGAAAGAGACACAGTGAAGAAAGCAGTGGTGGTCCCTGCGCGAGTGA	1706
Qy	1689	GACTTCTGGAGAGAGAGCCTGTGTAGAGCAGACACACTCTGGGCAATGACACAA	1748
Db	1707	AGCACTCGGGAGAGAGAGCCTTGTGTAGAGCAGACACCGCTGTGTGGGAGACGATCAAA	1766
Qy	1749	CTACAAATGCTGTGAAGCCAGAGAAAGCCAGAAAGCCACTGCTGGCCACCTTAAT	1808
Db	1767	TTACAAATGCAATTAAGCCAGAAAGACTG-----CTGCTCCCTGCGCTCACTGT	1817
Qy	1809	GAGTAAATGTATGATGATCAACCAAGAGCTGGCTTCCAGAGCCCTCCATCATCT----G	1863
Db	1818	GTAATAATGTATGTATGATCACTTAAGGGATTTGGCTCTCTGGAACCCCTCCGATCTTCTGAT	1877
Qy	1864	GGTGGCTGCTGGGAGCTGTGTAGACAGC----AGTCTGGGTGTTCTGTGCTGTATG	1918
Db	1878	AGCATTCCCTGGGAGCTGTGTAGAGCAGTGTGAGCTTGTGTGAAGCTGTATCAAG	1937
Qy	1919	TAGCAGCAGC-----TCACTGCCAGCAGAAAGCAGATACCAAGATGCTCTGAGACCCAG	1974
Db	1938	ACACCGGTGACCAATCAAGCGGGAGAGCAGACCCCAATGTCAGATGCTCTCTGCTGCTT	1997
Qy	1975	GTTGTTCTGCTAGACCTGTGTGAGCCTCTCTGGGTGTCTTAAAGCTGTGTAGCCTTC	2034
Db	1998	GTTGTTCTGCTAGACCTGTGTGAGCCTCTCTGTGTCACTGTTTCCAAAGCTGTAAACCTCAC	2057
Qy	2035	ATGT 2038	
Db	2058	TGTT 2061	
RESULT 10			
AA878224			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homio sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
REFERENCE			
1			
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saico,K., Yamamoto,J.,...			

Isihii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primers for synthesizing full-length cDNA and their use
JOURNAL Patent: JP 1074617-A 13129 07-FEB-2001;
Research Association for Biotechnology (JP)

FEATURES
Source Location/Qualifiers

1..2772
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
301..1989
/note="unassigned protein product"
/codon_start=1
/protein_id="CAE90061.1"
/db_xref="GI:40032961"

CDS

ADPLEPPPPQOQ/SLSTRSGKOPKTERVBOPLTIAARRKRSHPVSLDEGEPTSC
PATDERTASBSVESASSTRSGPQASATVAKRPASSEVKGADHDSDSDSLT
LKEI/LNRKRKEDEPTERPLKGIQSLRKRREBPATVSESDIVEGLPSKOB
PENIDGVVSOAGKDDRESLKBKAODIDEEPDLGRPEKIDYVPAALYCIKOP
HNHREMICCDRCBEMHFGCVGISBARGSLERNEDYI CPNCTILQVODETHSEAD
QOEAETRPDADGDTCTSIGTIEOKSSRDQGI KRI BKRAKNSPKKIKI PQVLEAP
GASKTIGCCCTVAPDSVYGCNDCLIKHAAATKATLSGSKQKRPKEMKMKPKP
SLPKGAOAGIKISSVHKRPAPKKRTVKA VVPARSEALGKACSSSTPSWASD
HNYNAVKERTAAPSPILYKCMYHGVGLDPSRSFWTAIPWACPGI/VNALC"

ORIGIN

Query Match 35.3%; Score 1013; DB 6; Length 2772;

Best Local Similarity 73.4%; Pred. No. 7e-300; Mismatches 475; Indels 36; Gaps 8;

Matches 1413; Conservative 0; Mismatches 475; Indels 36; Gaps 8;

141 GCTTGACCGCTGCTGACCTGACCCAGCTCATATTTATGATGATTTCCAGCCA 200
|||||
211 GCTTGACCCCAAGTTTGTCTGCAAGGTTTGTGTTATTAAGATTT-CAGGAA 269
201 AGTTTCCAGTTTCAAGTTTGGACAGATATGATGATTAAGGCACTGACATGA 260
270 AAGTGTCCAGCTTCAAGTTGAGCAGATGAGCAGCAAAAGCGCAGCAGATGA 329
261 GGAAGACCCCAAGGCTATTAACCCACAGTAAAGATTCAGGAAACCTGGGTTTGG 320
330 GAGGACACTTAAGGCATCAAAACCCACAGCAAAAGTTTCAAGAAACATGGGTTTGG 389
321 AAGAACCAAGATTGCAAAAGTGAAGGTGCAAGAGACAGAGGCGGACCCAGTAGCA 380
390 AAGGACCACTATTCGCAAGCAGAGGCGCAGAGGACGCGAGGCTGACCTGAGACC 449
381 GCAACC-----ACAGCAGCATTAACCTCTCCCTGCGCGCAGTGAACGCAACCAAAAG 434
450 GCCACCCCAAGCAGCAGCTGGGCTGTCCCTGCGGCGCAGTGGAGGCAAGCCCAAGCG 509
435 TACTGAGAGGTGAGAGATTTCTTACACAGTTCCGCGCCCGAGGAGAAAGATGTCC 494
510 CACTGAGCGGTGAGCAGATTTCTGACCATGTGCGCGCGCGCGGAGGAGGAGCATGTCC 569
495 GGTGTCCTGAGAGATTCAGTGAAGCCACATCTTCCAGAGTCACTGATGTGAGACAGC 554
570 TGTCTCTCGAGAGATTCGTGAGGCCACAGTCTCTCCCGCAGACGCGCAGAGCAGC 629
555 TTCCGAGGAGCGTTGAAAGCAGTTCTGAGATCAGAGTGGCCCTGTATCTGACTCTT 614
630 CTCGAGGCGCAGGTGAAAGCGCTTCTGAGACCAAGCGCGCCCAAGTCTGCTTCCAC 689
615 A---GGAAAGAACATCTCTCTCTTCTGAAAAGCAAAAGAGAGAGAGAGAGAAAG 671
690 ACTGTGAAGAGAACGACAGCTCTTCTGAAAAGGTGAAAGAGGAGAGAGAGAGAGAG 749
672 CACTCTGACAGTGAAGTGAAGCTTACAGTGAAGAGAACTTCAAGACCGCTTGGAG 721
750 CACTCTGACAGTGAAGTGAAGTGAAGCTTCAAGTGAAGAGCTTCAAGATGCTTGGCAG 809
732 AAAGCAGAGCAAGAACTGTGAGAGGTTCCTGAGAGGCAAGTCAAGATGCTTGGAGAA 791
810 GAAGCGGAGAAAGAGACCACTGAGAGGCGCCCTGAAAGGAGATCAAGATGCTGCGGAA 869

792 GAAGCCGAGAGAGAGAGATTTCTGCCGAACTGAGAGTGTCAAATAGCAGTCCAGCA 851
870 GAAGGCGCGGAGAGAGGATCCCGCGAGACTGTGGCTCCAGGCGCAGTACACTGAGGA 929
852 GACAGACCTCTCTGTAAAGCAGAGGCTGAGGCTATCAGGACCAAGTGTCCCAATCAGA 911
930 GGGCGCTCTGCCAGATGACAGAGGCCAGAGAACATCAGGGGGTGTGTCTCCAGGCTGG 989
912 GACAGATGACATTAAGAAATCAGTTGAGGAGAGAGGCGACTCAGGAGAAATCAGAGGAAA 971
990 GAAGATGACAGAGAGATTAAGTGAAGGAGAGAGGCGCTCAGAGACATCAAGATAGAGA 1049
972 CCCGAGGAGAGCGGCGAAACCAAGCCTGAGTGAAGTTTACGACCCCAATGCCCTGTGA 1031
1050 GCTTGAAGCTTGGCGCGACCGAGCTTAATGTAGGGTTAGACCCCAAGCCCTGTGA 1109
1032 CTGCATCTGCGCGACGCTTCAACAAAGTTTATGATCTGCTGTGATCGGTGAGGA 1091
1110 TTGCATTTGCCCGCAGCTTCAACAAAGTTTATGATTTGCTGTGACCGCTGTGAGGA 1169
1092 GTGATTCATGATGACTGTGTGGTATTTCTAGGCGCGAGGGGCGCTCCTGGAAGGAA 1151
1170 ATGATTTCAATGGGATGTGTGGCATTTCTGAGGCTCAGAGGAGGCTTTTGAAGAGAA 1229
1152 CGGAGAGACTACATCTGCCCAAAATGACCAATTTGCAAGTGCAGAGTGAACAACCG 1211
1230 TGGGAGAGACTATTTCTGCCCAATCTGCACATTTCTGCAAGTGCAGAGTGAAGCTCATTC 1289
1212 TTAGCGCACCAATGACAGGACTCTGGGTGACAGTCTGTGGTCTGATGAGCAGACTG 1271
1290 AGAAAGGAGATCAGCAGAGAGCTAAATGAGACCTGAGAGATGCTGATGACACCGATTG 1349
1272 CACAGACATAGGACAGTGAAGCAGAGTCCGAGAGACCAAGGCAATTAAGGTTAGAT 1331
1350 TACAAATATAGAGAAATATGAGCAGAGATCTACAGAAACCAAGGAGATTAAGGTTAGAT 1409
1332 TGAGAGGAGCAGCAAAACCCAGCGGCAAGAAAACTCAAGATATTCAGGCTGTGAGA 1391
1410 TGAAGAACTGCAAAATCCAAAGTGGCAAGAAAGAACTCAAGATCTTCAAGCTGTGATGA 1469
1392 GGGCTGTGGTCTCTAAATGACATTTGGCCCTGGGTTCAGATGACACAGCTGTGATC 1451
1470 GGGCGCTGGTCTCAAAATGATGTGGCCCGGGTCTGTCAAGTGGCGAGCGCGCATC 1529
1452 TGTGTATTGCAAGTAAAGATCTGATTTCTCAAAACGCGAGCAGCTACATGAGATTTCTAAG 1511
1530 GGTGTCTCAAGTAAATGACATTTCTCAAAACGCGCGAGCAGCAATGAGTTTCTAAG 1589
1512 TTCAAGTAAAGAACAAAAACAAAAACCAAGAGAAAGTTCAGACGACAGCAAAAGTT 1571
1590 CTGAGTAAAGAACAAAGCCAAAGCCTTAAGAAAGATGAATGAAGCAGAGAAAGCC 1649
1572 CAGTCTTCCAAATGACATTTCAAGTGGAGTTAAATCTTCTGTGCACAAGAGACT 1631
1650 CAGTCTTCCAAATGCGGTGTCTCAGGCAAGTATTAATCTTCTGTGCACAAGAGACC 1709
1632 AGGTGAGAGAAAGGAGAAACCAATGAGAA---AGTGAATCTGGCTTCCAGAGTGA 1688
1710 AGCTTCAAGAAAGAAAGAGACACAGTGAAGAGGACAGTGTGTCTCTCGGAGTGA 1769
1689 GACTTCTGAGAGAGGAGCCTGTGAGAGCAGACACATCTCTGGGCAAGTGAACCAAA 1748
1770 AGCACTCGGAGAGAGAGCTTGTGAGAGCAGACAGCGCTGTGGGCGAGAGATCAAA 1829
1749 CTACAAATGCTGTGAGCGCAGAGAGCCAGAGAAAGCCATGCACTGTGCCCAACCTATT 1808
1830 TTCAATATGCAATTAAGCAGAGAAAGACTG-----CTGCTCCCTGCGCGCTGACTGT 1880
1809 GACTTAATGTAACCTATTCACCAAGGCTGCTTCCAGGCGCCCTCCCATCATCT- - - - -G 1863
1881 GTATTAATGTATATATCACTTAAAGGTTGGCTCTGAGACCCCTCCCTCTTCTGTGAT 1940

QY	672	CACCTCGACAGTATGAGTATGCGCTTACCTGTAAAGAACTTCAGAAACGCCCTTCGAG	733
QY	673	CACCTCGACAGTATGAGTATGCGCTTACCTGTAAAGAACTTCAGAAACGCCCTTCGAG	734
Db	750	CACCTCGAGATGATACAGCAATGCGCTCGACCTTGAAGAAGCTTCAGAAATGCCCTTCGAG	809
QY	732	AAAGCTGAGCAGAACCTGTGAGAGGCTCCCTGAGAGCGATCAGAAATGCGCTTGAGAA	791
QY	810	GAAGCTGGGAAACAGAGGCCACTGAGAGGCCCTTAAAGAGGATCCAGATGCCCTGCGGAA	869
Db	792	GAAGCTGAGAGAGGAATTCGTGCGGAACCTGGGAGTGTCCAAATTAAGCAGGTGCCAGCA	851
QY	870	GAAGCTGCGGAGAGAGGCTCCCGCAGACCTGTGAGCTCCAGGCCCACTGACACTGTGGA	929
Db	852	GAAGCAGACTCTCTGTATAGCAGAGCCTGAGGCTATGTCAGGAGCAGATGTCCAGTCA	911
Db	930	GGGGTCTCTGCCAGTAAAGCAGAGGCCCGAGAACGATCAGGGGGTGTGTCTCCAGGCTGG	989
QY	912	GACGATGACATAGAAATCAGTTGGAAAGGAGGCCACTAGAGGAAATACAGAGGAAAA	971
Db	990	GAAGAATGACAGAGAGATGATTGAGAGGAAAGCGCTCAGACATCAAAAGATGAGAA	1049
QY	972	CCCCAGGGAAGCGGGCAAAACCAAGCGCTGAGTGTGAGGTTATACGACCCCAATGCCCTGA	1031
Db	1050	GCTTGAAAGACTTGGGCGCGAGCCGAAAGCTGAAATGTAGGGTTACAGCCCAAGCCCTGTA	1109
QY	1032	CTGCACTGTGCGCGCAGCGCTCAACAACAAGGTTATATGATCTGCTGTATGCGGTGAGAA	1091
Db	1110	TTGCACTTTGCGCGAGCCTCAACAACAAGGTTATATGATTTGCTGTACCGCTGTGAGAA	1169
QY	1092	GTGCTTCAATGATGATCTGTGTGGTATTTCTGAGAGCCCGAGGGGGCTCTGGAAGGAA	1151
Db	1170	ATGATTTCAATGCAATGTGTGTGGCATTTCTGAGGCTGAGAGGAGCTTTTGAAGAGAA	1229
QY	1152	CGGGGAAGACATCACTGTGCGCCCAAAATTGCAACATTTTGCAAGTGCAGATGAGCAAAACG	1211
Db	1230	TGGGGAGACATATATCTGCCCAACTGCACCATTTCTGCAATGTCAGAGATGAGACTATTC	1289
QY	1212	TAGCGCCACCAATGACAGAGCTCTGGGTGAGATCTGTGGGTGTGATGCGACAGACTG	1271
Db	1290	AGAAACCGCAGATCAGAGGAAGCTTAATGAGAGCTGTGAGATGTGATGACCGATTTG	1349
QY	1272	CACAGAGCTAGGAGCAGTATGAGAGCAAGACTCCGAGGAACACGAGGCATTAAGGGTAGAT	1331
Db	1350	TACAGATATAGGAACATATGAGCAGAAATCTTAGCGAAGACCAAGGATTAAGGGTAGAT	1409
QY	1332	TGAGAGGCAACAAACCCAGCGGCAAGAAAACTCAAGATATTCCAGCTGTCTGAGA	1391
Db	1410	TGAGAAAGCTCAAAATCCAAATGGGCAAGAAACTCAAGATCTTCAGCTGTGATAGA	1469
QY	1392	GGCTCTGTGTCTCTTAATATGATTTGGCTCTGGGTGTTCAGGTGTAGCAACGCTGACTC	1451
Db	1470	GGCGCTGTGTCTCTCAAAATGTATTATGGCCCGGGTGTGTGACGTGGGCGACGCCGACTC	1529
QY	1452	TGTGATATGAGTATGACTCATTTCAAAACAGCAGCAGCTACATGATATTCTTAAG	1511
Db	1530	GGTATCTGCAATATGACTGTATCTCAAAACCGCCGACGCAAAATGAAGTTCTTAAG	1589
QY	1512	TTCAAGTAAAGACAAAAAACAACCAAGGAAAGGTCAAGAGAAAGCAAAAGTT	1571
Db	1590	CTCAGGTAAAGACAAAGCCAAAGCCTTAAGAAAGATGAAGATGAAGCCAGAGAACCC	1649
QY	1572	CAGCTTCCAAAATGCAATGTTCAGGTGGGGTTAAATCTCTTCTGTGCAACAAGACT	1631
Db	1650	CAGCTTTCGAAATGCGGTGCTCAGGCAAGATTTAAATCTCTTCTGTGCAACAAGACC	1709
QY	1632	AGCGTCAAGAAAAAGGAAAAACCAAGTAAGA--AGTATGTCTGGCTTCAGAGTGA	1688
Db	1710	AGCTCAAGAAAAAAGAGACCAAGTAAAGAGGCAATGTGTGCTCCCTGCGCGGAGTGA	1765
QY	1689	GACTTCTGGGAGAGAGGAGCGCTGTGAGACAGCAACATCTGTGGGAAATGATCAAA	1748
Db	1770	AGCACTCGGAGAGAGGAGCGCTGTGAGACAGCAACGCGCTGTGTGGGAAAGATCAAA	1823

Qy	1749	CTACAAATGCTGTGAAGCCAGAGAAAGCCAGAGAAAGCCATGCACTCTGCGCCACACCTATT	1808
Db	1830	TTACAAATGCAATTAAGCCAGAAAAGACTG-----CTGCTCCCTCGCCGTCACTGTT	1880
Qy	1809	GAGTAAATGTACATATCAACCAAAAGCTGTGCTTCCAGGCCCTCCCATCATCT-----G	1863
Db	1881	GTAATAATGTATGTATCACTCAAGGGGTGTGGCTCTCTGGAACCCCTCCGTTCTTCTGGAT	1940
Qy	1864	GGTGGCTGCTCGGGGCTGTCTAGAACAG-----AGTCTTGGGTGTTCTGTGCTGTATAG	1918
Db	1941	AGGCATCCCTCGGGCTGTCCAGAGACTGGAGTGTGACGTTGTGTTAAGCTATATACAG	2000
Qy	1919	TAGCCAGACG----TCACTGCCAGCCCAAGAAAGATACCAAGATGCTCTGGAACCCAG	1974
Db	2001	ACACCGGCTGTACCCATCAGGGGGAAGAGAGCCATCTCAGAAATGCTCTCTGCTGCCCT	2066
Qy	1975	GTTGTTCCGCTAGCCTGTGTGAAGCTCTCTGTGGGTGTTCTTAAAGACTGTGTAGGCTCTC	2034
Db	2061	GTGTCCATTCCTAGTCTGTCAAGAGACTTCTCTCACTGTTTTCAAAGCTGTAAACCTCAC	2120
Qy	2035	ATGT 2038	
Db	2121	TGGT 2124	

RESULT 13
BC000770

DATE 13-NOV-2003
-PWT 145000

[illegible]

ACCESSION	BC000770
VERSION	BC000770.2
KEYWORDS	GI:33990873
	MGC.

SOURCE ORGANISM	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
Mammalia, Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2407)
AUTHORS
Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altshul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Ussid, T. B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullany, S. J., Bosak, S. A., McEwan, P. J., McKenna, K. J., Majek, J. A., Gunaratne, P. H., Richards, S.

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Heltton, B., Kettman, M., Madan, A., Rodrigues, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Boufard, G.G., Blakesley, R.W., Touchman, J.W., Green, B.D.,
Dickson M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.B., Scherch, A., Schein, J.B., Jones, S.U. and Marra, M.A.: Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MENTURE
22388267

PUBMED
 12477932
 2 (bases 1 to 2407)
 Straube B
 11/10/03

AUTHOR TITLE JOURNAL DIRECT SUBMISSION Submitted (15-NOV-2000) Gene Collection (MGC)	National Institutes of Health, Mammalian Cancer Genomics Office, National Cancer Institute
---	--

DEMANDY
 nru-wc@project.mri. <http://mac.nci.nih.gov>
 USA
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 Some Collaboration Project, International

<p>KARPERNA</p> <p>MAIL ROOM ONLY - RECEIVING On Aug 20, 2003 this sequence version replaced gi:12653952.</p> <p>COMMENT:</p> <p>Contact: MGC help desk Email: cgaarb-remail.nih.gov</p>	<p>KARPERNA</p> <p>MAIL ROOM ONLY - RECEIVING On Aug 20, 2003 this sequence version replaced gi:12653952.</p> <p>COMMENT:</p> <p>Contact: MGC help desk Email: cgaarb-remail.nih.gov</p>
--	--

Tissue Procurement: DCTD/BTP

CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
http://www.systembiology.org
contact: amadan@systembiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://Image.lnl.gov
Series: IRAL Plate: 7 Row: 1 Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

FEATURES

source
1. .2407
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:3257 IMAGB:3506207"
/issue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"
/note="Vector: pOT87"
1. .2407
/gene="DATP1"
/note="synonyms: D101, FLJ11265, KIAA0333"
/db_xref="locusid:11083"
/db_xref="MIM:604140"
201. .1835
/codon_start=1
/product="death associated transcription factor 1, isoform b"
/protein_id="AAH00770.1"
/db_xref="GI:12653953"
/db_xref="LocusID:11083"
/translation="MDKSDPSNVEAPKAIKPTSKFRKTWGRRTTIAREGAGDAE
APLPPPOOLGLSLRSGOPKRTVEBOFLTAARRGSMVSLSDSEPTSC
PATDAPTAESGVSASSTRSPQASATVRRPASREKVGCDHDDPSDPLT
LRLQRLNRKKEQEPTEPLKGIQSLRKRRREBPATVSGEADTVBGVLPSPQE
PNDQVVSQAGDQRESRLRGAQDIDBEPGDGRPECEGVDPAALYCTCRQP
HNRPMICDRCEWPHGDCVGISSEARGLERNGEDYICPNCTIIQVDDTSEPAD
QDAKVRPGDADCTSIGTIEOKSSBOGIKRIEKAAPSGKKLKIPOVIRAP
GASKTGGCCVADPSVYCNDCILKHAATMFKLSGKKKPKPKKMKPKPK
SLPKCAQGIKISVHRRPAPRKETTYKAVVVPARBEALKEKAAACSSSTPSMSAD
HNYNAVERKTAAPSPSLYKSGKLYSLHPSLIA"

misc_feature
1011. .1160
/note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="CD:pfam00628"

ORIGIN

Query Match 34.4%; Score 985; DB 9; Length 2407;
Best Local Similarity 76.3%; Pred. No. 2.9e-291;
Matches 1266; Conservative 0; Mismatches 380; Indels 13; Gaps 4;

Db 141 GCTGGCGCTGCTGACCTGCGCCAGGCTATATTTATATGATTCACGCA 200
111 GCTTGACCCAGAGGTTGCTGCGCCAGGCTTTTGTGTATTTAGGATTT-CAGGAA 169
Qy 201 AGTTTCAAGCTTTCAGTTTGGGACAGGTATGATTAAGGACACTGACATTA 260
Db 170 AGGTGTCAGGCTTTCAGTTTGGGACAGGTATGATTAAGGACACTGACATTA 229
Qy 261 GGAAGCAACCAAGCTATCAAAACCAACAGTAAGATTCAGGAAAACCTGCGTTTTCG 320
Db 230 GAGAGCACTTAAGGACATCAAAACCAACAGTAAGATTCAGGAAAACCTGCGTTTTCG 289
Qy 321 AAGAACCAAGATTCGCAACGTAAGGCTGAGAGACAGAGGCGAGCCCACTGAGACA 380
Db 290 AAGGACCACTATCGCCAGAGAGGCGCAGGAGGCGGAGGCTGACCCACTGAGAGCC 349

Qy 381 GCAACC-----ACAGCAGCATTAACCTCTCCCTGCGCCGAGTGAAGCAACCAACG 434
Db 350 GCCACCCCAACAGCAGCAGCTGCGGCTCTGCTCGCGCGAGTGGAGGAGCCCAAGCG 409
Qy 435 TACTGAGAGGTGAGAGATTTCTTACACGCTTCCGCGCCGAGGAGAAAAGATGCTCC 494
Db 410 CACTGAGCGGTGAGCAGTCTTCTGACCATTTGGCGCGCGCGCGCGAGAGAGCAATCC 469
Qy 495 GGTGCTCCCTGAGAGATTCAGTGAAGCCCATCTTCCAGATCACTGATGTGAGACAGC 554
Db 470 TGTCTCCCTGAGAGATTTCTGTAAGCCCATGCTCTGCGCCAGAGCCGAGAGCAGC 529
Qy 555 TTCGAGAGGAGGCTTGAAGAGAGCTTCTGATCAGAGTGGCCCTGATCTGATCTCTT 614
Db 530 CTCGAGGCGAGGTGAGAAAGCCCTTCTGAGACAGAAAGGCGGCCCAAGCTGCTTCAC 589
Qy 615 AGG--GAAAGACATCTCTCTCTTCTGAAAGGCAAAAGAGTGAAGAGAGAA 671
Db 590 AGCTGTGAAGAGACAGACGACCTCTTCTGAAAGGTGAAGAGGAGATGACCAAGATGA 649
Qy 672 CACTCTGACAGTGAAGATGATGAGCTTACGTTGAAGGAACCTTCAAGAACGCTTGAG 731
Db 650 CACTCTGATGATGACAGGATGAGCTTACCTTGAAGAGACCTTCAAGATGACCTTCCAG 709
Qy 732 AAGCAGAGCAAGAACTGTGAGAGGTCCTGAGAGGCACTGAGATCGCTTGAAGAA 791
Db 710 GAAGCGGAGAACAGAGCCCACTGAGAGGCGCTGAAAGGATCCAGATGCGCTTCCGAG 769
Qy 792 GAAGCGAGAGAGAGATTTCTCCCAAACTGGAAGTGTCCAAATAGAGTGCAGTCCAGCA 851
Db 770 GAAGCGCGGAGAGAGGCTCCCGCAGACGTGAGGCTCCGAGGCGAGTCACTGTGAG 829
Qy 852 GGAAGACCTCTGTGAAGAGAGGCTGAGGCTAGCAGGAGCAAGTGTCCAGTCA 911
Db 830 GGGCGTCTGCGCCAGTGAAGAGAGGCCAGAAAGATCAAGAGGATGTGTCCAGGCTGG 889
Qy 912 GACAGATGACATTAAGAAATCAATTGAGAGGAGAGGCACTCAGGAGAAATCAGAGAGAA 971
Db 890 GAAGATGACAGAGAGATGATGAGAGGAAAGCGCGCTCAGAGACATCAAGATGAGAG 949
Qy 972 CCCGAGGAGAGCGGAGCAAAAGCCTGATGTGAGATTTAGACCCCAATGCTCTGTA 1031
Db 950 GCTGAGAGATTGAGGCGAGCCAGAGCCTGAATGTGAGGTTAGACCCCAAGCCTGTGA 1009
Qy 1032 CTGCATCTGCTGCGCAGCTTCAACCAAGCTTTATGATCTGCTGATGATGAGTGTGAG 1091
Db 1010 TTGCATTTGCGCCAGCTTCAACCAAGCTTTATGATTTGCTGTGACGCTGTGAGAA 1069
Qy 1092 GTGATTCATGATGATGATGATGATTTCTGAGAGCCGAGGCGGCTCTGAGAAAGAA 1151
Db 1070 ATGATTTCAATGAGATTTGTGTGGGCAATTTCTGAGGCTCAGAGGAGAGCTTTTGA 1129
Qy 1152 CGGAGAGAGCTACATCTGCGCAAAATTCACATTTTTCAGAGTGAAGATGAGCAAAAG 1211
Db 1130 TGGGAGAGAGCTATATCTGCGCAAACTGACATTTCTGCAAGTCAAGATGAGACTATTC 1189
Qy 1212 TAGCGCAACCAATGAGCAGAGCTCTGGGTCAGATCTGTGTGGGCTGATGAGCAAGCTG 1271
Db 1190 AGAAAGGAGAGATCAGCAGAGAGCTAAATGAGAACCTGAGAGATGCTGATGAGCAAGCTG 1249
Qy 1272 CACAGCATTAAGGAGCAGTGAAGAGAGTCCGAGAGAGACAGAGGCAATAAGGATAGAT 1331
Db 1250 TACAGATTAAGAGCAATGAGAGAGAGTCTGACGAGAGCAAGGATTAAGGATAGAT 1309
Qy 1332 TGAAGAGGAGCAAAACCCAGCGCAAGAAATTCAGAGATTTCCAGCTCTGCTGAGA 1391
Db 1310 TGAAGAGGAGCAAAATTCAGAGTGAAGAGAACTCAAGATCTTCAAGCTCTGATGAGA 1369
Qy 1392 GGTCTCTGATGCTCTCAAAATGATTTGGGCTGGGCTGTCAAGTGTGACAGGCTGATCTC 1451
Db 1370 GGTGCTGATGCTCTCAAAATGATTTGGGCTGGGCTGTCAAGTGTGACAGGCTGATCTC 1429
Qy 1452 TGTGATTCAGATTAAGTCAATCTCAAAACAGCAGAGGCTTACATGAGATTTCTTAAG 1511


```

Db      1430 GGTGACTGCGATATGACTGATCTTCTCAACACCGCGAGCGACATGAACTTCTTAG 1489
Qy      1512 TTCAAGTAAAGACAAACAAACCAAGGAAAGGTCAAGACGACGAAAGTT 1571
Db      1490 CTCAGGTAAAGACAAAGCCAAACCTTAAAGAAAGATGAAGTGAAGCCAGAGAGCC 1549
Qy      1572 CAGTCTTCCAAATGACAGTGTTCAGGTGGGATTTAAATCTCTTCTGTCACAAAGACT 1631
Db      1550 CAGTCTTCCAAATGACAGTGTTCAGGTGGGATTTAAATCTCTTCTGTCACAAAGACT 1609
Qy      1632 AGCGTCAGAGAAAGGAAAGCAAGCCAGTGAAGAA---AGTATGTCGCTTCAGAGACTA 1688
Db      1610 AGCTCCAGAGAAAGGAAAGCAAGCCAGTGAAGAAAGGAGTGTCTCCCTGCGGAGACTA 1669
Qy      1689 GACTTCTGGAGAGAGGAGCGCTGTGAGAGCAGACACACATCTGCGCAAGTACACAA 1748
Db      1670 AGCACTCGGAGAGAGAGCGCTGTGAGAGCAGACACGCGTGTGGGCGAGGATCACA 1729
Qy      1749 CTACAACTGCTGAGAGCCAGAGACCCAGAGAGCCAC 1787
Db      1730 TTACAATGACATGAAGCCAGAGAAAGATGCTGCTCCTC 1768

RESULT 14
LOCUS   AC119716      221547 bp    DNA       linear   HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
ACCESSION AC119716
VERSION    AC119716.6 GI:25092366
KEYWORDS   HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus
          1 (bases 1 to 221547)
REFERENCE 1
AUTHORS   Muzny,D,Marie, Metzker,M,Lee, Abramson,S, Adams,C, Alder,J,
          Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
          Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
          Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benhamed,F,
          Biewald,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
          Bryant,N, Buhay,C, Burch,P, Cavazos,I, Ceasar,H, Calderon,B,
          Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Calderon,B,
          Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
          Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
          Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
          Delgado,O, Denson,S, Derramo,C, Ding,Y, Dinh,H, Divya,K,
          Drepper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
          Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
          Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
          Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
          Gebregorjais,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,M,
          Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K,
          Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
          Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogue,M,
          Hollins,B, Howells,L, Huiyk,S, Hume,J, Idlebird,D, Jackson,A,
          Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
          Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
          Kowalski,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
          Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
          Lotenshew,L, Louisedge,H, Lozado,R,J, Lu,X, Ma,U,
          Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A,
          Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,B,
          Manwiley,S, McLeod,M,P, McNeill,T,Z, Neenen,B,
          Mliostavljevic,A, Miner,G, Minja,B, Montemayor,J, Moore,S,
          Morgan,M, Morris,K, Morris,S, Munilasa,M, Murphy,M, Nair,L,
          Namberyis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
          Nwakeremem,O, Okunonu,G, Olarinmunesoon,A, Pal,S, Parks,K,
          Pasternak,S, Paul,H, Perez,L, Perez,L, Pfannkuch,C,
          Plopper,F, Polinder,A, Popovic,D, Primus,B, Pu,L,-L,
          Puzos,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,

```

```

          Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
          Rivers,C, Rodkey,T, Rojase,A, Rose,M, Rose,R, Ruiz,S,J,
          Sanders,W, Savery,G, Scherer,S, Scott,C, Shattman,S, Shen,H,
          Shetty,J, Shwartzbeyn,A, Sisson,I, Sitter,C,D, Smejs,D,
          Sneed,A, Sodergren,B, Song,X,-Z, Sorelle,R, Sosa,J,
          Steidle,M, Strong,R, Sutton,A, Svatek,A, Tabors,P, Taylor,C,
          Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K,
          Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J,
          Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
          Williams,G, Willson,R, Wlarczyk,R, Wooden,H, Worley,K,
          Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V,
          Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von
          Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O,
          Weinstock,G and Gibbs,R.A.
          Direct Submission
          Unpublished
          2 (bases 1 to 221547)
          Worley,K.C.
          Direct Submission
          Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          3 (bases 1 to 221547)
          Rat Genome Sequencing Consortium.
          Direct Submission
          Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Nov 19, 2002 this sequence version replaced gi:23611286.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          ----- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          ----- Project Information
          Center project name: GWLC
          Center clone name: CH230-274116
          ----- Summary Statistics
          Assembly program: Phrap; version 0.990329
          Consensus quality: 208013 bases at least Q40
          Consensus quality: 211685 bases at least Q30
          Consensus quality: 213907 bases at least Q20
          Estimated insert size: 199182; sum-of-contigs estimation
          Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 17 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          * 1 198201: contig of 198201 bp in length
          * 198202 198301: gap of unknown length
          * 198302 198749: contig of 1448 bp in length
          * 198750 199849: gap of unknown length
          * 199850 201115: contig of 1266 bp in length
          * 201116 201215: gap of unknown length

```

* 201216 202338: contig of 1123 bp in length
* 202439 202438: gap of unknown length
* 202439 203500: contig of 1062 bp in length
* 203501 203500: gap of unknown length
* 203601 204505: contig of 1305 bp in length
* 204906 205005: gap of unknown length
* 205006 206192: contig of 1187 bp in length
* 206193 206292: gap of unknown length
* 206293 207385: contig of 1093 bp in length
* 207386 207485: gap of unknown length
* 207486 208519: contig of 1034 bp in length
* 208520 208619: gap of unknown length
* 208620 209758: contig of 1139 bp in length
* 209759 209858: gap of unknown length
* 209859 211441: contig of 1583 bp in length
* 211442 211541: gap of unknown length
* 211542 212876: contig of 1335 bp in length
* 212877 212976: gap of unknown length
* 212977 214256: contig of 1280 bp in length
* 214257 214356: gap of unknown length
* 214357 217003: contig of 2647 bp in length
* 217004 217103: gap of unknown length
* 217104 218459: contig of 1356 bp in length
* 218460 218559: gap of unknown length
* 218560 219788: contig of 1229 bp in length
* 219789 219888: gap of unknown length
* 219889 221547: contig of 1659 bp in length.

FEATURES
source
1. 221547
Location/Qualifiers

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-274116"
1. 1272
/note="wgs_end_extension
clone_end:Sp6"
complement(4416..5263)
/note="clone boundary
clone_end:Sp6
site:
end_sequence:32266352"
misc_feature
192445..193553
/note="wgs contig"
195743..196592
/note="clone boundary
clone_end:T7
site:
end_sequence:32266349"

ORIGIN

Query Match 29.8%; Score 853.4; DB 2; Length 221547;

Best Local Similarity 81.7%; Pred. No. 2.1e-250; Indels 112; Gaps 1;
Matches 979; Conservative 0; Mismatches 107;

53 TTAAGAGCAGTCCCACTGTCCTTGGTGTTCGAAGCTCCGGAATCTTCTATTGAT 112
Db TTAAGAGCAGTCCCACTGTCCTTGGTGTTCGAAGCTCCGGAATCTTCTATTGAT 206532
113 GAACGTGACGTGACAAACAGTGGGGTGAAGCTTGGCCGTCTGCTTGAACCTGACCGCCAGGTC 172
Db GAACGTGACGTGACAAACAGTGGGGTGAAGCTTGGCCGTCTGCTTGAACCTGACCGCCAGGTC 206412
173 TATAATTTTATAGATTTTCCAGCCAAAGTTTCCAAAGCTTTCAGTTTGGACGAGC----- 227
Db TATAATTTTATAGATTTTCCAGCCAAAGTTTCCAAAGCTTTCAGTTTGGACGAGC 206472
228 ----- 227
Db TTTTGTTTTATGCAAGTCTCCCTCAAGATGAACATGCAATTAACCTTGTCCATTACTA 206532
228 -----AGTATGATGAT 240
Db GCAATTAATGCAATTAAGCCTGACAGCCAGAAAGAAATTTCAAGGTATGATGAT 206592

QY 241 AAAGGCACTTGAGCAATGAGAGCAACCCAAAGCTATCAAAACCAAGTAAGAGTTTC 300
Db AAAGGCACTTGAGCAATGAGAGCAACCCAAAGCTATCAAAACCAAGTAAGAGTTTC 206552
QY 301 AGAAAACTTGGGGTTTTCGAAGAACCAAGATTTGCCAAACGTGAAGGTTCAGAGACAGC 360
Db AGAAAACTTGGGGTTTTCGAAGAACCAAGATTTGCCAAACGTGAAGGTTCAGAGACAGC 206712
QY 361 GAGGGGACCCCAAGTGAAGCAACCAAGAGCAATTAACCTTCCCTGGCCGCAAGTGA 420
Db GAGGGGACCCCAAGTGAAGCAACCAAGAGCAATTAACCTTCCCTGGCCGCAAGTGA 206772
QY 421 CCGCAACCAAAAGTATGAGAGGGTGAAGAGTTCTTACCAAGGTTCCGCGCCAGAGG 480
Db CCGCAACCAAAAGTATGAGAGGGTGAAGAGTTCTTACCAAGGTTCCGCGCCAGAGG 206832
QY 481 AAAAAGATGTCCCGGTGTCCCTGAGATTTCCAGTGAAGCCCAATCTTCCAGATCACT 540
Db AAAAAGATGTCCCGGTGTCCCTGAGATTTCCAGTGAAGCCCAATCTTCCAGATCACT 206892
QY 541 GATGCGAGACAGCTTCCGAGGGGAGCCCTGAAGCAAGTTCTGAAGTCAAGATGCGCT 600
Db GATGCGAGACAGCTTCCGAGGGGAGCCCTGAAGCAAGTTCTGAAGTCAAGATGCGCT 206952
QY 601 GTATGACTCTCTTGAAGAAACATCTGCTCTTCTGAAGAAAGCAAAAGAGGTGA 660
Db GTATGACTCTCTTGAAGAAACATCTGCTCTTCTGAAGAAAGCAAAAGAGGTGA 207012
QY 661 GAGGAGAAAGCACTTGAAGATGAGTGAATGCTTACCTTGAAGAACTTCAAGAC 720
Db GAGGAGAAAGCACTTGAAGATGAGTGAATGCTTACCTTGAAGAACTTCAAGAC 207072
QY 721 CCCTTCCGAGAAAGGAGAGCAAGAACTGTGAGAGGTCCTGAGAGCAAGTCAAT 780
Db CCCTTCCGAGAAAGGAGAGCAAGAACTGTGAGAGGTCCTGAGAGCAAGTCAAT 207132
QY 781 CCCTGAGAGAAAGGAGAGCAAGAACTGTGAGAGGTCCTGAGAGTTCGAATAGC 840
Db CCCTGAGAGAAAGGAGAGCAAGAACTGTGAGAGGTCCTGAGAGTTCGAATAGC 207192
QY 841 AGTCCGAGCAAGACAGACCTTCTGTAGCAAGAGCTGAGGCTTACAGGACCAAGT 900
Db AGTCCGAGCAAGACAGACCTTCTGTAGCAAGAGCTGAGGCTTACAGGACCAAGT 207252
QY 901 TCCCACTGAGACAGATGACATGAAATCAAGTTGGAAGGAGGCACTCAGGAAAT 960
Db TCCCACTGAGACAGATGACATGAAATCAAGTTGGAAGGAGGCACTCAGGAAAT 207312
QY 961 ACAGAGAAACCCCAAGGAGACGGGCAACCAAGCCTGAGTGAAGTTTACGACCC 1020
Db ACAGAGAAACCCCAAGGAGACGGGCAACCAAGCCTGAGTGAAGTTTACGACCC 207372
QY 1021 AATGCCCTGACTGATCTGCGCCAGCCTCAACAACAGGTTATGATCTGCTGAT 1080
Db AATGCCCTGACTGATCTGCGCCAGCCTCAACAACAGGTTATGATCTGCTGAT 207432
QY 1081 CCGTGTGAGAGTGTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1138
Db CCGTGTGAGAGTGTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 207490
RESULT 15
AC108337 272404 bp DNA linear HTG 21-SEP-2002
LOCUS Rattus norvegicus clone CH230-115113, *** SEQUENCING IN PROGRESS
DEFINITION *** 2 unordered pieces.
AC108337
AC108337.4 GI:23266147
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 272404)

Munry,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Albrooks,S, Amin,A, Angiano,D, Ayala-Bechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Bielawski,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Burch,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, Dedetich,D, Davila,M, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinn,H, Divya,K, Drexler,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,B, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M, Hollins,B, Howells,S, Huily,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,S, Jolivet,A, Kapachy,S, Kelly,S, Khan,Z, King,L, Kovar,K, Kowar,K, Kowalski,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshewa,L, Louised,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mahoney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,D, Moore,S, Morgan,M, Morris,K, Morris,S, Mundaasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwokilemech,O, Okwuonu,G, Olarnpungsoon,A, Pal,S, Parks,K, Paternacki,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C, Plopper,P, Poindexter,A, Popovic,D, Prims,E, Pu,L, Puzo,M, Qutroz,J, Rachelin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Rutz,S,J, Sanders,W, Savary,G, Scherer,S, Scott,G, Shatman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smaj,D, Sneed,A, Sodergren,E, Song,X,Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Uemami,K, Valas,R, Vera,V, Villalana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willison,R, Wiczysk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Zhou,D, von Weinstock,G, and Gibbs,R.A.

TITLE
JOURNAL

Direct Submission

2 (bases 1 to 272404)

REFERENCE
AUTHORS

Worley,K,C

TITLE
JOURNAL

Direct Submission

Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

3 (bases 1 to 272404)

TITLE
JOURNAL

Rat Genome Sequencing Consortium.

REFERENCE
AUTHORS

Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Sep 21, 2002 this sequence version replaced gi:21736648. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Acllas (<http://www.hgsc.bcm.tmc.edu/projects/rt/>). As a result, the contigs may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only

contigs will be indicated in the feature table.

----- Genome Center of Medicine

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GPMX

Center clone name: CH230-115113

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 238513 bases at least Q40

Consensus quality: 241910 bases at least Q30

Consensus quality: 244206 bases at least Q20

Estimated insert size: 262154; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 270622: contig of 270622 bp in length

* 270623 270722: gap of unknown length

* 270723 272404: contig of 1682 bp in length.

----- Location/Qualifiers

1. 272404

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-115113"

1. 1256

/note="wgs end extension

clone_end:5p6"

complement(1814..2663)

/note="clone boundary

clone_end:5p6"

site:BCORI

end_sequence:BH294125"

62809..64046

/note="wgs contig"

267959..268006

/note="clone boundary

clone_end:77

site:BCORI

end_sequence:BH294091"

ORIGIN

Query Match 26.1%; Score 748.2; DB 2; Length 272404;

Best Local Similarity 92.1%; Pred. No. 5.8e-218;

Matches 789; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 228 AGGTATGATGATTAAGGACCTGAGCAATGAGAGCAACCAAGGCTATCAACCCAC 128338

Db 128279 AGGTATGATGATTAAGGACCTGAGCAATGAGAGCAACCAAGGCTATCAACCCAC 128338

QY 288 CAGTAAAGGTTTCAAGAAACCTGAGGTTTGAAGAAACCAAGTTCACCAAGGAGG 347

Db 128339 CAGTAAAGGTTTCAAGAAACCTGAGGTTTGAAGAAACCAAGTTCACCAAGGAGG 128338

QY 348 TGCAGAGACACCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 407

Db 128339 TGCAGAGACACCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128458

QY 408 GCGCCGAGTGAACGCGCAACCAACGTAAGAGGTTTCTTAACCAAGGT 467

Db 128459 GCGCCGAGTGAACGCGCAACCAACGTAAGAGGTTTCTTAACCAAGGT 128518

```
OY      468 TCAGCGCCGAGGAGAAAGAAATGTCCCGGTGTCCCTGAGAGATTCCAGTGAAGCCCAATC 527
Db      128519 TCAGCGCCGAGGAGAAAGAAATGTCCCGGTGTCCCTGAGAGATTCCAGTGAAGCCCAATC 128578
OY      528 TTCACAGTCACTGATGTGAGACAGCTTCCGAGGGAGCCGTGAAGCAAGTTCTGAGAT 587
Db      128579 CTCACAGTCACTGATGTGAGACAGCTTCCGAGGGAGCCGTGAAGCAAGTTCTGAGAT 128638
OY      588 CAGAAAGTGGCCCTGATCTGACTCTTGAAGGAAAGACATCCCTGCTTTCTGAAGAGGC 647
Db      128639 CAGAAAGTGGCCCTGATCTGACTCTTGAAGGAAAGACATCCCTGCTTTCTGAAGAGGT 128698
OY      648 AAAAGAGGTGAAGAGAGAGAGACACCTCTGACAGTGAACAGTATGACCTTACGTTGAA 707
Db      128699 GAAAGAGGTGAAGAGAGAGAGACACCTCTGACAGTGAACAGTATGACCTTACGTTGAA 128758
OY      708 GGAACCTTGAAACCCCTTCCGAGAAAGCGAGAGCAAGAACCTGTGAGAGAGTCCCTGAG 767
Db      128759 GGAACCTTGAAACCCCTTCCGAGAAAGCGAGAGCAAGAACCTGTGAGAGAGTCCCTGAG 128818
OY      768 AGGCGATGAGATGCGCTGAGAGAGAGCGCAGAGAGAAATTCGCGAAACTGAGAG 827
Db      128819 AGGCGATGAGATGCGCTGAGAGAGAGCGCGAGAGAGAAATTCGCGAAAGCTGAGAG 128878
OY      828 TGTCCAATTAGCAGTGCAGAGCAGAGACGACCTCTCTGTGAAGCAGAGACCTGAGGCTAG 887
Db      128879 TGTCCAATTAGCAGTGCAGAGAGAGAGACGACCTCTCTGTGAAGCAGAGACCTGAGGCTAG 128938
OY      888 TCAGGAGCCAGTGTCCCACTCAGAGACAGATGACATAGAAATCACTTGAAGGAGAGGC 947
Db      128939 CCAGGAGCCAGTGTCCCACTCAGAGAGAGATGACAGAGAAATCACTTGAAGGAGAGGC 128998
OY      948 GACTCAGGAGAAATACAGAGAGAAACCCGAGGAGAGCGGCAAAACCAAGCTGAGGTGA 1007
Db      128999 GACTCAGGAGAAATACAGAGAGAAACCCGAGGAGAGCGGCAAAACCAAGCTGAGGTGA 129058
OY      1008 GGTTTACGACCCCAATGACCCTGTACTGATCTGCGCCAGGCTCACAAACAAGTTTAT 1067
Db      129059 GGTTTACGACCCCAATGACCCTGTACTGATCTGCGCCAGGCTCACAAACAAGTTTAT 129118
OY      1068 GATCTGCTGTGATCGGT 1084
Db      129119 GGAGGAGGGAACCTGGT 129135
```

Search completed: April 23, 2004, 14:16:35
Job time : 10938.8 secs